

```

FT REPEAT 2778 2781 6.
FT REPEAT 2782 2785 7.
SQ SEQUENCE 2867 AA; 331433 MW; 6E7D8CA71AFBFFD3 CRC64;

Query Match 4.3%; Score 231; DB 1; Length 2867;
Best Local Similarity 18.9%; Pred. No. 0.0055;
Matches 243; Conservative 219; Mismatches 436; Indels 388; Gaps 65;

QY 11 LALTTVAVTYSGVGLERESVKGQQTOS---ASEDMFEEDNRKTNVSGNSTV- 65
DB 1314 IATTKGTSPTSQDINEL---ESIKEVHNKLQLVKQSSNMEW---MRKQILSKMDLIL 1368
QY 66 --DETVEIDLPSDG-----NSNNSSKTESVVS-----DEKQVPAKPEVTOASN 108
DB 1369 NNSFTIAKEISNNTQNALGFRENKTKTANKTDELQVAAAMIEBAKHKNNDIALDAQ 1428
QY 109 SSNDASKEVPEKQDTASKKETTLETSTWBAKDFTRGDTLVGFSQSGINKLSQTSHTLPS 168
DB 1429 IDTEVSKIEQINREIMNKQKIEKSYLSIKEYKDKCTEISNSKRGDKIEFLE----- 1482
QY 169 HAADGTQLTVASAFAPDPK-----KTAIAEYTSRLBENGKPSRLDID--OKETIDGRI 221
DB 1483 -----KFPNESSNSKNVNNINEINNT--RNSEQYIKDIEDAKQASTVEL 1527
QY 222 FNAYQTLTLTPNGYKSGDPAFYDNKNIAEVNLPESLETISDYAFAMSLKQYKLPDNL 281
DB 1528 FHKHE---TTISNFKK--SEILGVETSKQKINAEDIKIEHNSSEIQTVAGPQENL 1583
QY 282 KVIQELAFPDQIGKLYLPHILIKLAERAKSNRIQVEVLESLKLYIGASAPD--NN 339
DB 1584 NKLMEPHNYDN-----ADELNNKK-----STNAKVLLETNLESYKHN 1621
QY 340 LRNV-LPDGLKIES-----EAFNGPDE-----HYNNQVVLRT 376
DB 1622 LSETNKGGEKLYSKAKDMOKIKASENTAEKTEKVDQSNVYNIQT---T 1677
QY 377 QGNPHQATE-----NTYVNDPKSL--WRATPDMDYTKLE-----DFTYQK 417
DB 1678 ERN--LIVTEKNRNLNGIDSTITNEGALKESKNGVEIFLEKLEIGNRKLKVDITKKS 1735
QY 418 -NSTVG-----FSNKGLQKVRNKL-----EIPQHNQITTEIGDPAFRNV----- 459
DB 1736 INSTVGSSLENNFDLNQYFNKNINDYENKMGIVYEPFG--SLNKISEN--LKNASENT 1793
QY 460 -DPOS-KTLR-----KYDL-----EIKLPTIRKICAPAFQSNLKSFASEDLERI- 505
DB 1794 SDYNSAKTLRLBAQEKVNLANKREANKIADYKVSFRP-----IFNMESLDKIN 1847
QY 506 ---KEGAFMNRNIGTL-DLQDKLIKIGDAAFHNHIAVILP-----ESVQEIQRSAFR 555
DB 1848 EMIKKEQLTVNEGHNVKQVLVENIKELVDE---NNLSDLIKQATGKNEIQKITHSTLK 1903
QY 556 QNGALHLMFIGNKVTIG-----EMAFISNGLSVNLSQKQK-----TIEQASD 603
DB 1904 NKAKTILIHVDTSARYVGKITPELA-LTELLGAKLTAAQELKESKNNVLETEWSK 1962
QY 604 NALSEVVLPEMLQ-----TIREBAFK-----RNLH 628
DB 1963 NT-NELDVHKNIQDAVYVALETILANSDELDTQKQSSKLIMGNQIYLVKVLINQYKKI 2021
QY 629 KEVKS-----STLSQITF-----NAFDQ---N 648
DB 2022 SSIKREBAVSVKIGVNSKSELSKITCSQDSYNNIATLEKQTELQMLRNSFTQEKNT 2081
QY 649 DGDGRFGK-----KYVVRT-----HNSHM-----LADGERFIID 678
DB 2082 NSDSLKLEKITDPSLKALKTLEBENALKAASSNHEVQSKSEPVNPALSEIKETD 2141
QY 679 PDKLSSTWDLKLVKILIEGLDYSTRLOTTOFQEMTTA-----GKALLS--KSNLR 729
DB 2142 IDSMTALDELKKGRTGVSNYKLIKDTVTEKISDDDELINTIEKNVAYLAIVKQVYE 2201
QY 730 QGEKQKFLQEAQFLGRV-----DLDKAIABAKALVTKATKNGHLIENSINKAVILAY 783

```

```

DB 2202 DTQVQVLTINHEHNTQVSNHEPTNFKSNKSSSE--LTKAVTDSKTIILSKL--KGVILE 2257
QY 784 NNSAIKANKVRLKLEKLDLITDLVEKGPLAQTMWQGYVLTKPLPLPEYVIGLWVPD 843
DB 2258 VNNTMTNNTTSSAKIEALVYELNKK-----KTSINLEYQSNVEKLDKMSNAKYID 2312
QY 844 KSGKLTYALDMSPTIEGGQKAVGNPILNVEDNEGHTLAVATLADYEGLYIKDILNNS 903
DB 2313 VS-----KIFPTVLDTQK---SNIVTQHSINN-----YKDKLKGK 2345
QY 904 LDK-IRAIROIPLAKYHRLG-IFQAIRMAABADRLPKTPKYLVNENYRKKQKQKNTL 961
DB 2346 LQELIDADSSFTLESTIKFENBIYSHIKTNIGLEHQ-LQQTNKSEHNVAHGKEKIVH--- 2401
QY 962 KPYDYKTPINAKLPHKVDGDAAGH-----YINMETNSVAVTPIRSE 1007
DB 2402 -----LINRV---ESLKGD--VKNHDDQYMKKLNASLNDNINKTNSINISDEELK 2448
QY 1008 QQLHKSQSDVNLPTQSSKNFIYEIL 1033
DB 2449 KLLKQVENDQLCKNNNTQNFISDIM 2474

RESULT 2
NUM1 YEAST STANDARD; PRT; 2748 AA.
AC 000402; 003767;
DT 01-DEC-1892 (Rel. 24, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Nuclear migration protein NUM1.
OS NUM1 OR YDR150W OR YD8358.06.
GN Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 28383 / Fli100;
RX MEDLINE=92079907; PubMed=1745235;
RA Kormanec J., Schaff-Gerstenschlaeger I., Zimmermann F.K.,
RA Percecho D., Kuentzel H.;
RT "Nuclear migration in Saccharomyces cerevisiae is controlled by the
RL highly repetitive 313 kDa NUM1 protein.";
RL Mol. Genet. 230:277-287(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1678;
RX MEDLINE=97313263; PubMed=9169867;
RA Jacq C., Alt-Moerbe J., Andre B., Arnold W., Bahr A., Ballesta J.P.G.,
RA Batgeys M., Baron L., Becker A., Biteau N., Bioecker H., Blugeon C.,
RA Bokovic J., Brandt P., Brueckner M., Buittago M.J., Coster F.,
RA Delavaue T., del Rey F., Dujon B., Fide L.G., Garcia-Cantalejo J.M.,
RA Goffeau A., Gomez-Perris A., Granotier C., Hanemann V., Hannein T.,
RA Hohnleil J.D., Jaeger W., Jimenez A., Doniaux J.-L., Kraemer C.,
RA Kuster H., Laamanen P., Legros Y., Louis E.J., Moeller-Rieker S.,
RA Monnet A., Moro M., Mueller-Auer S., Nussbumer B., Paricio N.,
RA Paulin L., Pera J., Perez-Alonso M., Perez-Ortin J.E., Pohl T.M.,
RA Prydz H., Purnelle B., Raemussen S.W., Remacha M., Revuelta J.L.,
RA Rigter M., Salom D., Saluz H.P., Saiz J.E., Soren A.-M., Schaefer M.,
RA Scharte M., Schmidt E.R., Schneider C., Scholler P., Schwarz S.,
RA Uristetazau L.A., Verhasselt P., Vlesers S., Voet M., Volckaert G.,
RA Wagner G., Wambuit R., Wedler E., Wedler H., Woelfl S., Harris D.E.,
RA Bowman S., Brown D., Churcher C.M., Connor R., Dedman K., Gentles S.,
RA Hanlin N., Hunt S., Jones L., McDonald S., Murphy L., Niblett D.,
RA Odell C., Oliver K., Rajandream M.A., Richards C., Shore L.,
RA Walsh S.V., Barrell B.G., Dietrich F.S., Mulligan J.T., Allen E.,
RA Arayo R., Aviles E., Berno A., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Hunnicke-Smith S., Hyman R.W., Komp C.,
RA Laaskari D., Lew H., Lin D., Mosedale D., Nakahara K., Namach A.,
RA Oefner P., Oh C., Petel F.X., Roberts D., Schramm S., Schroeder M.,
RA Shogren T., Shroff N., Winant A., Yelton M.A., Bolstein D.,

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd

OM protein - protein search, using sw model

Run on: April 27, 2004, 11:10:26 ; Search time 17 Seconds

3231.413 Million cell updates/sec

Sequence: 1 MTKKHLKTLALITVSVT.....VSLCLFLVTAGKKGRARK 1055

scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	231	4.3	2867	1	RBP2_PLAVB	000799 plasmodium
2	224	4.2	2748	1	NM1_YEAST	000402 saccharomy
3	215.5	4.0	1727	1	ALM1_SCHPO	000351 schizosacc
4	209	3.9	1208	1	PCP1_SCHPO	0923t5 schizosacc
5	209	3.9	1875	1	MLP1_YEAST	002455 saccharomy
6	208	3.9	1013	1	SCA4_RICRH	002455 saccharomy
7	203	3.8	998	1	SCA4_RICAK	09a181 rickettsia
8	202.5	3.8	2869	1	RBP1_PLAVB	000799 plasmodium
9	202	3.8	1022	1	SCA4_RICPR	000799 plasmodium
10	200	3.7	939	1	TE95_TETTH	09z449 rickettsia
11	200	3.7	939	1	IF2_WOLSU	07m755 wolnelliella
12	199.5	3.7	3259	1	GOB1_HUMAN	014789 homo sapiens
13	199	3.7	1357	1	KTN1_HUMAN	08b6u2 homo sapiens
14	199	3.7	1957	1	SPOF_SCHPO	010411 echinosacc
15	198.5	3.7	1790	1	USO1_YEAST	P25366 saccharomy
16	198	3.7	1010	1	SCA4_RICPA	09a175 rickettsia
17	196.5	3.7	992	1	SCA4_RICRY	09a175 rickettsia
18	195.5	3.7	1969	1	MYSA_CAREL	P12844 caenorhabd
19	192.5	3.6	1358	1	SIR4_YEAST	R11978 saccharomy
20	192.5	3.6	1839	1	CYAA_SACKL	P23466 saccharomy
21	192	3.6	1022	1	SCA4_RICCN	P25268 rickettsia
22	189.5	3.5	961	1	SCA4_RICFL	09a137 rickettsia
23	188.5	3.5	1012	1	SCA4_RICSL	09a180 rickettsia
24	188	3.5	1011	1	SCA4_RICAF	09a183 rickettsia
25	187	3.5	1928	1	MYSL_YEAST	P08964 saccharomy
26	186.5	3.5	1011	1	SCA4_RICMN	09a182 rickettsia
27	185	3.5	1018	1	SCA4_RICTA	09a179 rickettsia
28	195	3.5	1972	1	MYHB_PABIT	P35748 oryctolagus
29	184.5	3.4	1539	1	Y373_HUMAN	015078 homo sapiens
30	183.5	3.4	1315	1	CHAO_DROME	P12024 drosophila
31	183.5	3.4	8545	1	CANCI_CAREL	09n444 caenorhabd
32	183	3.4	2653	1	CENE_HUMAN	000224 homo sapiens
33	181	3.4	991	1	SCA4_RICSI	09a177 rickettsia

35	181	3-4	1334	1	SMC4 SCHEP	P14004	schizosacch
34	181	3-4	1690	1	C190 DROME	09v155	drospothia
36	180.5	3-4	1979	1	TRTA HUMAN	015433	home sapien
37	179.5	3-4	1163	1	SRCT CLOAB	027451	home sapien
38	179.5	3-4	1972	1	MYHB_HUMAN	935749	home sapien
39	179	3-3	1310	1	RASO YEAST	1212753	saccharomyc
40	179	3-3	1320	1	CENF_HUMAN	P494544	home sapien
41	178.5	3-3	1159	1	EX5B BORBU	0151788	borrellia bu
42	178.5	3-3	1658	1	Y657 YEAST	Q03661	saccharomyc
43	177.5	3-3	1701	1	MSB1 PLAAF	013819	plasmodium
44	177.5	3-3	1972	1	MYHB_MOUSE	P08658	mus musculum
45	177	3-3	1637	1	MRSP_STAUU	P80544	staphylococc

ALIGNMENTS

ID	RBP2_PLAVB	STANDARD;	PRT; 2867 AA.
AC	000759; GONZM3;		
DT	01-APR-1993 (Rel. 25, Created)		
DT	15-MAR-2004 (Rel. 43, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Retículoocyte binding protein 2 precursor (PVRBP-2).		
GN	RBP-2 OR RBP2.		
OS	Plasmodium vivax (strain Belen).		
CC	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.		
OX	NCBI_TaxID=31273;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.		
RX	MEDLINE=20299192; PubMed=10838223;		
RA	Galinaki M.R., Xu W., Barnwell J.W.;		
RT	"Plasmodium vivax reticuloocyte binding protein-2 (PvRBP-2) shares structural features with PvRBP-1 and the Plasmodium yoelii 235 kDa rhoptry protein family.";		
RL	Woj. Biochem. Parasitol. 108:257-262(2000).		
RN	[2]		
RP	SEQUENCE OF 1189-2439 FROM N.A.		
RX	MEDLINE=9231538; PubMed=1617731;		
RA	Galinaki M.R., Medina C.C., Ingravallo P., Barnwell J.W.;		
RT	"A reticuloocyte-binding protein complex of Plasmodium vivax merozoites."		
RL	Cell 69:1213-1226(1992).		
CC	-1- FUNCTION: Involved in reticuloocyte adhesion. Specifically binds to human reticuloocyte cells.		
CC	-1- SUBCELLULAR LOCATION: Membrane-bound (Probable).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/submit/ or send an email to license@ebi.ac.uk).		
CC	-----		
DR	EMBL; AF184623; AAF76525.1; -.		
DR	HSSP; P03069; IGCM.		
DM	Malaria; Receptor; Signal; Transmembrane; Repeat.		
FT	SIGNAL	21	POTENTIAL.
FT	CHAIN	22 2867	RETICULOCTE BINDING PROTEIN 2.
FT	DOMAIN	22 2805	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	2805 2826	POTENTIAL.
FT	DOMAIN	2827 2867	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	44 133	ASN-RICH.
FT	DOMAIN	560 758	LYS-RICH.
FT	DOMAIN	1112 1285	LYS-RICH.
FT	DOMAIN	2758 2785	7 X 4 AA TANDEM REPEATS OF H-D-D-T.
FT	REPEAT	2758 2761	1.
FT	REPEAT	2762 2765	2.
FT	REPEAT	2766 2769	3.
FT	REPEAT	2770 2773	4.
FT	REPEAT	2774 2777	5.

QY 991 I 991
Db 1231 V 1231

RESULT 3
ALM1 SCHPO STANDARD: PRT: 1727 AA.
AC Q9UTK5: 013313: Q9UTK8:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Abnormal long morphology protein 1 (Sp8).
OS ALM1 OR SPAC1486.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles M., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson K.,
RA Holroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens K.,
RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Wolke R.G., Aert R., Robben J., Grymopoulos B.,
RA Welter J., Vanstele E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Bozzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler R., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Lucas M., Rochet M., Gallard C., Hunt C., Moore K., Hurst S.M.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Uesery D., Barrell B.G., Nuree P.,
RA "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880 (2002).
RN [2]
RP SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.
RC STRAIN=972;
RX MEDLINE=20123449; PubMed=10660053;
RA Jimenez M., Petit T., Canedo C., Goday C.,
RA "The alm1 gene from Schizosaccharomyces pombe encodes a coiled-coil
RT protein that associates with the medial region during mitosis.";
RT Mol. Gen. Genet. 262:921-930 (2000).
RN [3]
RP SEQUENCE OF 644-834 FROM N.A.
RC STRAIN=968 h90;
RX MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA Hiraoka Y.,
RT "Large-scale screening of intracellular protein localization in living
RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RT Genes Cells 5:169-190 (2000).
CC -1- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
CC CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
CC CYTOKINESIS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
CC or send an email to license@ebi.ac.uk).

DR EMBL: AL133357; CAB2414.1;
DR EMBL: AF010473; AAB55416.1; ALT INIT.
DR EMBL: AB028012; BAA87316.1;
DR PIR: T50073; T50073.
DR GenBank: SPombe; SPAC1486.04C;
KW Coiled coil.
FT DOMAIN 57 361 COILED COIL (POTENTIAL).
FT 443 463 COILED COIL (POTENTIAL).
FT DOMAIN 542 740 COILED COIL (POTENTIAL).
FT 804 1106 COILED COIL (POTENTIAL).
FT DOMAIN 1223 1427 COILED COIL (POTENTIAL).
FT 1497 1555 COILED COIL (POTENTIAL).
FT DOMAIN 1601 1664 COILED COIL (POTENTIAL).
SQ SEQUENCE 1727 AA; 197858 MM; F820HFD9C132644 CRC64;
Query Match 4.0%; Score 215.5; DB 1; Length 1727;
Best Local Similarity 19.3%; Pred. No. 0.015;
Matches 231; Conservative 198; Mismatches 422; Indels 343; Gaps 57;
QY 12 ALTVSVVTVTSGEYVGLEREESVQKQGTQASD-----DWFEEDNERKTVNSKNSGTVD 66
Db 116 ALTT-----EWQSIRANSELQEQSKIASBQLIAKQIDALQENSHLGRQVQSAH 167
QY 67 ETVSD-----LFSGNNNSKTESVSPDKQVPKAKPEVTQASNSNDASVYEV 118
Db 168 QALSDIERKKQKHFA-----SSSKVKEELIVQ-----KSAVSDIASLSQSDHSKV-- 215
QY 119 PKDPTAKSKETLETSTWEADPVTGDTLVGFSKSG--INKLSQTSHLVLPASHADG--- 173
Db 216 -----CKLEFVSRLQVQDLEKK--LAGLAQNTLENLEKQIFPKRKSNSVSDNIS 264
QY 174 -----TQITQVASFPTDCKTALAEYTSRGENGKSPRLDID---QKEITDEGRIF 222
Db 265 KILETDPTSTIKELKEEVEYTKRLTALME-----SKSELQSEVALAQEKLTQCSLY 316
QY 223 NAVQATK-----LTINGYKSIQD--DAFVNDKNIABVN-----LPE 257
Db 317 N--NVYELANNQOQLLISNSIRLEQKTDVSVSEQVYKKNKNTSVSAGVLFSPLAQ 374
QY 258 SLETTIS--DYAFAMSLIKQVLPDNL-----KVIGELAFPDNQTGKLYLPRHLIKL 307
Db 375 KLSAVQNPERSFTKVSQDNMKLQKVSLKQLDRLTNKFSFCEQV--KQRIIP--VVKQ 430
QY 308 AERAFKSNRIQYVEFLGSKLVIGASFDNNIRNVMLPDGLEKISEAFATGNGDBHYN 367
Db 431 QSEIYKNNIY--NWFLESSELET-----SNNNLTQVQELSTKMQEAC-----YL 475
QY 368 NQVVLRTQTQNPQD---LATENTYVNPDKSLRAPPDMDYTKLREDFYQKNSVYG-- 422
Db 476 QLASHTQCSDSLREYVLCMAELDHLNETKS--RNPVATQVA-LDE--YAQNPSTASE 529
QY 423 -FSNKGIQ-----KVRNKNLEIPKQNGITITIGDNAFRNVDFOSKTLKRYDLEIKL 476
Db 530 TLVKNKELANPSSIKEAVSKTLEIRKVRALEC-----DVEIQKQTVQVQISNAVKE 580
QY 477 PSYTRIKGAPAPQSNNLKSTFABEDLEIKGAFNNR-----IG 516
Db 581 NSNT-----LSBOIKNISELSNSKIKNESILNERNLKMLATSRSSILSHNSAG 632
QY 517 TLIDKDKLKIGDAFHINIYVILVPE--SVQEIGRSAPRQNGALHMPGKVKVTIGE 574
Db 633 NID--DKMSIDSTLEKNYVYNNEMTAID--SLSKRQD-----LSE 676
QY 575 MAPLSNKLSEVNSSEKQKLT-----IEVQAFSNALSEVVLPPNLTQTI--REAFKR 625
Db 677 MEAIRKELE--NSKYQOOLSTDRLTNANDVDEAFKKEAKELRSINQLODIISRQDQRAS 724

QY 626 NHTKXGSSSTLSTITFNAFQNDGKRFKKVYVTRNNHMLADGERFIIDPKLST 685
 DQ 735 KFAEELHVNLSARKEKLNASKEGDLKRT-----QERLISENDK--- 778
 QY 686 WDLKFKVKTIEGLDYSTRQTOTOTOFREMTTAGKALLSKNLROGEKOFLOAOFELG 745
 DQ 779 LAERERLMSIVSD-----OTFLNQ--QOLSDAR-----KVFESKE--- 815
 QY 746 RVDLDKAKAKAKALVTKKAKNKHLLERSINKAVLAVNN-----SAIKKANYKRL 796
 DQ 816 --SLSLSLQKKEE--NEKSNNDLHSLQKSLKSGIEYSRIKTMLEKOSLSBDNRKLL 871
 QY 797 E-----KELDLTDVVEGKGPLAOATWQ-----GYLLKTPILPEYITGL 838
 DQ 872 DNOGMEIKYQELNGVIELEKORSTIEAKTQOKRSTYSERREALLSSLSJDSKRTSL 931
 QY 839 NVTFDKSGKLIALDMSDTIGEGQ---KDAYGNPIINVDNEGYHTLAVATLADYEG 894
 DQ 932 ESQVNSILRNIEQLAASKLAEBWVETVEYRLOTSLSLEKN--LMTSLSEKRLVI 990
 QY 895 YIDLIINSLSDKIKAIQIPLAKHRLGIRQALNAAEAD---RLPKTPKGTALNEVPN 951
 DQ 991 LQDEIASSSL-----RCENITKDETRVALLIENKHLNNELSS 1029
 QY 952 YRKQMEKNLKPVDYKTPI-----FNKALPN 977
 DQ 1030 HRNAEKQHEKENDYKQQLLLVTEDEKTEDEKELLRHADARSTIQKREDYTTALQ 1089
 QY 978 -EKVDGDRPAK-GHINIAETNNSVAVTPISSEQQLHKSQSDVNI.POTSSKN-NF 1028
 DQ 1090 VEDLNKEIALKAGINESQPPISKEKEDPLROEVYVLLKQNAMLTLQSSNMF 1143

RESULT 4
 PCPI_SCHPO STANDARD; PRT; 1208 AA.
 ID PCPI_SCHPO
 AC 092351;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Spindle pole body protein pcpi.
 GN PCPI OR SPAC69.06C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RE SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
 RX MEDLINE=21852775; PubMed=11864908;
 RA Flory M.R., Morpheu M., Joseph J.D., Means A.R., Davis T.N.,
 RT "Pcpi, a Spcl10p-related calmodulin target at the centrosome of the
 RL fission yeast Schizosaccharomyces pombe.";
 RL Cell Growth Differ. 13:47-58(2002).
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgourou J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,
 RA Brooks K., Brown R., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Gentes S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Moorey P., Moutle S., Mungall K., Murphy L., Niblett D., Ocell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmons M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymopoulos B.,
 RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Meller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Botzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Punelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Laure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Roover M., Gallatin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Pothakhin J.,
 RA Sipakovski G.V., Ussery D., Barrell B.G., Nure P.;
 RA "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -FUNCTION: Spindle pole body component that binds calmodulin.
 CC Overexpression of pcpi causes the formation of supernumerary SPB-
 CC like structures and disrupts both mitotic spindle assembly and
 CC chromosome segregation.
 CC -SUBCELLULAR LOCATION: Spindle pole body.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; Z81317; CAB03608.1; -
 CC EMBL; AF348506; AAK31344.1; -
 CC PIR; T39068; T39068.
 CC GeneDB Spombe; SPAC69.06C; -
 CC Calmodulin-binding; Coiled coil.
 FT DOMAIN 151 375
 FT DOMAIN 387 803 COILED COIL (POTENTIAL).
 FT DOMAIN 874 1091 COILED COIL (POTENTIAL).
 FT DOMAIN 1177 1204 COILED COIL (POTENTIAL).
 FT SEQUENCE 1208 AA; 140763 MW; 70264159AD62424 CRC64;

Query Match 3.9%; Score 209; DB 1; Length 1208;
 Best Local Similarity 19.4%; Pred. No. 0.018;
 Matches 220; Conservative 170; Mismatches 367; Indels 378; Gaps 52;

QY 20 TVSQEYGLERESYVQEOGTGSA-----SEDDMFEDNRKNTVSGNSTVDETVSDLS 74
 DQ 93 SDPNSTYGL---SAISKQATQERLSTISQNDST---DVSKLTDLSK-NSRIHTDDEL-- 143
 QY 75 DQNSNNSSKTESYVSD-----PKQVPA---KPEVTQASN 108
 DQ 144 PANPALTLREQKVKLEKVSRENFGRIKIVCLEKRLSMAAPQIKNAVVDNVEIHAERAN 203
 QY 109 SSSNDASKVAVPPQDTRASKKETLETSTWEANDVTRNGDTLVGFSKSGINLSQ-----T 161
 DQ 204 IQLQKRTESLLQKEDKNFKLE-----EKVDYLSK-----VNDVDSQNVKVF 248
 QY 162 SHLYVPSHADGQTQTVNA-SFAFPPDKKTALAVY-----TSRLGNGKSRRLDIDOK 213
 DQ 249 ERIRPLENALKVQKQKSLSTEMEEDKSNKKEVDYETELRQLONRIDELSE--ELDVAQ- 305
 QY 214 EIIDGELFNAYQTKLTIPNGYKSIQDAFYDNKNIAEVNPE----- 257
 DQ 306 DILTEKE-----DEIATLKQIIEKENSSAFENENSSVYHIOEDYALIIQAKODEPADRI 361
 QY 258 -----SLERTISYAAHNSLKQVLPDNLKIVGELAFPNQIGKLYLPRHLIKAEAP 312
 DQ 362 QVLTADLEKEKNQIMHSEASIGLTDNQV-----H--TLQEOULH 400
 QY 313 KSNRIQVTEFLOSKLKVGE--ASFQDNNTLRVMTLPGLEKTESAFTNGPDGHEHNNOV 370
 DQ 401 KAN--EIELEFHDQSRNNEBKKNEDIMLQSRSLSEEBDVLESKIQT---LEDDNNSL 454
 QY 371 VLRIT- GQNHQALTEBNTYVNPDKSLWPAIPMDYTKMLBEDFTYQKNSVTGFSNKGIQ 429
 DQ 455 RLMTSLSQIQLSIFATQRIIDEERKHLRLASKNSDALAE-----TNIRIQ 502


```

0Y 551 R$AFRQNALHMFIGNVKTIGEMAFPLSNKLESYNLESEQOKTIEVIOASDNALSEV 610
Db 611 -----KTSVPENBASYN-VTIKOLETIKRDLDESQVO----- 641
0Y 611 LPENIOTTIREAFKGNHLEKVGSSSTLSQIFPNAPDONDGDRFGKAVVTHNNSHMLA 670
Db 642 ---DIQT-----RISQITRETBMSILNKELIQUIDYBSKSISITKLGEKASSRILA 689
0Y 671 DGERPIIDPDKLSSTMDVLKLYKLIIEGIDYSTLRQTTQO--FREMTAGKALLS----- 724
Db 690 E-ERFKLSTNLTLITKANEDQLRKRPDYLONTILKQDSKTHELTNEVYSC-KSKLSIYET 747
0Y 725 -----KSNLR-----QCEKQKFIQEOAFPIGR 746
Db 748 ELNLNKEBQKLRVHLEKNLKOELNLSPEKOSLRIMVOTOLTOLOKREDDLBEET-----R 802
0Y 747 VDLDRALNAKEKAL-VTKKAT--KNNGH--LERSINKAVLAYNNNAIKKANVRLKEELD 801
Db 803 KSCQCKTIDELBALSELKKEYSQDHHIKOLEEDNNSIETWYON-----KIEALKOYE 856
0Y 802 LLTDLVEBKGPLAQATWQVYLLKTPLEPPEYIIGLVNVPDKSGKLIYALDMSPTIGEG 861
Db 857 SVITSVDSK---QDIEKLTQYKVS-----LEK----- 881
0Y 862 QNDAYAGNFIENVDENEGYHTLAV--ATLADYEGLYIOLINSSIDTK-----AIROI 914
Db 882 -----BLEEBEKIRLHTYWNDEIIND-----DSLRKELEKSKINULTAYSOIK 924
0Y 915 LAKYHRLGIPOAIRNAALAEADRLLEKTEKGYLNEVPYRKKOMENKULPVDYKPIEN-- 972
Db 925 EYK---DLYETTSOSLQOTNSKIDESTFDQFNQJKNLTDEKTSLEDDXISILKEOMFNLN 980
0Y 973 -----KALPNEKVDGDRRAKGNHINAEYNNAVATEVIRS--EOLIKSOSDVN---LP 1020
Db 981 NELDLOKKGMEKEKADPFK-----RISLONNNKEVAVKSPYBESKLSKINDDQOITY 1035
0Y 1021 QTS$KNNPIYEL 1032
Db 1036 ANTAQNNEYOEI 1047

```

```

RESULT 6
SCA4_RICRH
ID _SCA4_RICRH          STANDARD:          PRT: 1013 AA.
AC 09A081;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (P8120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia rhipicephali.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OC NCBI_TaxID=33992;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RT 'gene D' coding for an intracytoplasmic protein."
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
CC 1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF155053; AAK30684.1; -

```

QY	Antigen	1	1	3.9%	Score 208	DB 1	Length 1013
FT	NON TER	1013	1013				
FT	NON TER	1013	1013				
SO	SEQUENCE	1013 AA	110550 MW	856898912315D102	CRCE4		
	Query Match			3.9%	Score 208	DB 1	Length 1013
	Best Local Similarity			19.4%	Pred. No. 0.016		
	Matches	194	Conservative	148	Mismatches	322	Indels 336
							Gaps 47
QY	53	ERKTNVSKENSTVDETVDLFPDGDGNSNNS	-----SSKTESV-VSPDKCY----	95			
DB	108	EQKREIREBEKOKTISTFF	-GNPAREPFDKALENPELKKEKLEIAGKVNHNFF	165			
QY	96	-----PKAKBETVQASNSNDASKVBPVKODTASKEKTELTSTWEANDF	-VTRGD-TL	147			
DB	166	STAGYPGGFFKPVQWENQVSAADL	-RAVTVKNDADDELCITNETTWTKTPFTVAKDGTQ	224			
QY	148	VGFSK-----SGNKLSTQSHLPLPSHADGQTQTVASPAFPDDKTAI	-----AAYS	197			
DB	225	VQISSYREIDPPIKLDKPADGSMHLSMAVKKADGK	-----PSKDAIYFTAHYEE	274			
QY	198	RLGNGKPSRLDQKEIIDBGELEFNAYQLTKLTTPNGYKSGIDGAFVDNKNIAEYNLDE		257			
DB	275	-----GPGKKP-----		280			
QY	258	SLETTSDYAFAMSLQKVKLPNKLKVG	-----ELAFPNQIGSKL	-----LPHLLITLAER	310		
DB	281	-----OLKEISSPPPLKFAGTGDDAIAYIEH	-----GGELTYTLAVTRGKYEMMK	325			
QY	311	AFKSNRIQTVAFPLGSKLVIGASFPQDNL	-----RNWMLPDQLEKIESEAFQNG	362			
DB	326	EVELANGQSVDSQAEIITIGQSGKEQPLTPQQTASSVSP	PQYKQVPPPTPTNP	384			
QY	363	DEHYNQVVLPRTRQGNPHQATENTYVNPDKSLMRATPMDYTKLLEDPFYQKNSVTG	422				
DB	385	-----LQPETSQPMQSQOVNPNLTAATALGSMQDL	-----LN	418			
QY	423	FSNKGLOK-VKRNKMLEIPKO	-----NNG	-----ITTEIGDNARN	458		
DB	419	YVNAQLTEITISNKQIIDLKEAATLILHNASDIARQNTIATLMENTYVNNONLTPDAKY	478				
QY	459	-----VDFQSKTKR	-----VDLEBKK-LPSTTRKIGAFQFQSNLKSFEASEDLBEIKEGATM	511			
DB	479	AGVAAVLETINKNDQMTPLEKSMQLEATV	-----ATTINSENLBPQKEQMLEBTVD	530			
QY	512	NNRIGTDLQKCLKIGDAPAHINHIYAVLPESVQEIFGRSAFRONGALHMFIGNKVTY	571				
DB	531	-----VG-LSLKQD	-----ASRAAAIDGITDAVIKSNS	-----TEDKGTMLTAVGQKV	573		
QY	572	IGWAFLSNKLKESVNLSEOKOL	-----KTIEQQAQSDNALSEVNLPPNLTQTRBEERFKN	626			
DB	574	-----NVSELSNAERKOKLQSVLKKGVSEQAQVLS	-----PAQQLMOQNLDKITTAQOTKQ	623			
QY	627	HLKEVKG-----SST	-----LSQITFNAFP	-----ON	648		
DB	624	TIKKVNDILFPLPLSTELKTNINQAITSNVLDGPATAEVKEBIIQETINTTVAGSSLEAD		683			
QY	649	DGDKRFQKGVYVTRH	-----NNSHMLADGRFTID	-----PKD	-----LSSTMVD	688	
DB	684	KAEIYVKGGEITATIHSDTSLPNKALITMASEKIVGSKTNLPRELTMTKGLVIGIYEG		743			
QY	689	-----LEKVLKITEGLDSTYLRQTPQFQPREMTTAGKAL	-----LSKSNLROGEKQKFOE	739			
DB	744	KGGEPIITAASVSGIDNSININSEKALKKAKDABEATLIDITQNTLBEIKQGNIBEHKP		803			
QY	740	-----AQFLGRVD	-----LDK	-----AIAKAEKALVTYKATKNGHLERSIN	779		
DB	804	RDDIYNKQEVYINAVNPVTEALBEKPAKPVSAEERIYQETSILNNISGLAVEKVNFPRA		863			
QY	780	VLAIVNNAIKRAANKYRLKEKLDLTDLYEGSGPLAQATMVQGVYLLKTPPLPEYITIGLN	839				
DB	864	MLSSNGN	-----FTLEKKKEESITKYDELIVKAFG	-----TKSTEEQQSFITKN	908		

QY 840 VYFDKSGKLIYALDMSPTI-----GEGQKDAYGNPIIANDV 875
 DB 909 LIDDKTSLSKERVRLQITDKLQEOOTQKQAEIENPVSXTED 948

RESULT 7
 SCA4_RICAK STANDARD; PRT; 998 AA.
 ID SCA4_RICAK
 AC 09AIX9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
 GN SCA4 OR D.
 OS Rickettsia akari.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OK NCBI_taxid=786;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sekeyova Z., Roux V., Raoult D.;
 RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
 RT 'gene D' coding for an intracytoplasmic protein";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC
 CC EMBL; AF213016; AAK30691.1; -
 DR Antigen.
 FT NON_TER 1
 FT NON_TER 998
 SQ SEQUENCE 998 AA; 109328 MW; FCEB43AC62DE5BD5 CRC64;

Query Match 3.8%; Score 203; DB 1; Length 998;
 Best local similarity 18.9%; Pred. No. 0.027;
 Matches 220; Conservative 176; Mismatches 405; Indels 360; Gaps 55;

QY 6 LKTLAALTTSVVTYSQVYGLERESVKEQQTQASAE-----DMFEEDNERKT 56
 DB 57 ISTLSGTISTDDQI--SDPTKAVAREIITIOQKDEIAEQILKDLAALVDRDLAEQKKEI 114
 QY 57 NVSKENSTVDETVSLFSDGNSNNSSKTESVSPKQVPRKAPR-----VT 103
 DB 115 EEEKKEK--DKKLSVVF--GNPAN-----REFIDALEKPELKKLSEIETGYNNILIT 164
 QY 104 QEASN-----SENDASKVEVPKODTASKEETLETSTWEAKDFVTGRDTLY 148
 DB 165 YSAANGHGFGFPVQWENQISASDL-RATVYVNDAGDELCTINETVTWKPTVA----- 218
 QY 149 GFSKSG-----INKLSQTSILVLPASHADGTQLTQVASFAPTPKXTAI 192
 DB 219 --KKGTQVQINSYRAIDFPILDKRADGSMHLSMALXKADGTK-----PEKDAV 266
 QY 193 ---AAYTSRLGNGKPSRLDIOKEIIDEGELFNAYQLTKLTPNGYSIQGDAAVNDKN 249
 DB 267 YFTAHYEE--GPNGR----- 280
 QY 250 IAEVNI,PSLEFETISDYAFAMSLKQVYKLDNLKVIQE---LAFEDNOIGKLY--LPR 302
 DB 281 -----QUKRISSQPKPKFAGDDGDAVAIYEH--GGEITYTLAVTR 317
 QY 303 HLIIKLAERAFKSNRIQTV---EPLGSKL-KYVIGESAFQNNIRNVMMLPGLGK--IESNA 356
 DB 318 GKTKEMMEKVELHOGOSVDLSQI IAEIDLTKVGRS---QETLIQPIITPQELKSSIEPT 374

QY 357 FTGNPDGEHNNQVYLRLTRGTQNPQOLATENTYVNPDKSLMRATPDM-----DYTKLE 411
 DB 375 TTQVPPITPASQPV--HTISQMP-----QSQVNP--NLFPAATLSCSMQDLNLYVA 425
 QY 412 DFTYQK--NSVTGFSNKGLOKVRBNKLEIPIKONGIITTE--IGDNF-----RVNDF 461
 DB 426 GLTKEDQNTQGLDINEAATLILNN---EKEQANETITLKNVNNALTPDKVARVA 482
 QY 462 QSKTLR-KYDLEIKLPSTIRIKIGAFAPQSNLIKSEASDELEIE----- 507
 DB 483 VLETTIKNODPTPIEKSMLAEATVAITLNSENLTQKQKQOMTEKAVDVLSPFDTSRAV 542
 QY 508 -----GAPMNRIGLIDLDKLIKIGDAFHNHLYALVPSVQOIGSARONGALH 561
 DB 543 AIDGITGAIVKSNLSTKDGTMILAVGDYV-----NASELSNAEQ----- 584
 QY 562 IMFIGNKVTIGEMAFSLNKLSESVNLSEQKQLTIEVQAFSDNALSEVVLPPNLIQITRES 621
 DB 585 --LLGSVLKKGVETKILS-----PEQQQL-----MQQMLDKITAE 617
 QY 622 AFKRNHLKEVKGSSITLSQITFNAFDQNDGKREKQVYVTRHNSHMLD--GERFIIDP 679
 DB 618 QTRNDNITEVQG--ILANPAFNTIATAIQLKTTKVL-----DSPITAEIKGE----- 664
 QY 680 DKLSSTWVDLEKYLITIEGHVSTLRQTQTQFREMTTAGKALLSKSNLRQGEKQKFLQE 739
 DB 665 -----TLESTIKIVA--ESPINVOQKDTI--VKGMEALASHRTMAPTKKIAIE- 710
 QY 740 AQEFLGRVLDKAIARAEKALVTKKATKNGHLIERSINKAVLYNNSAIKKANVGRLEKE 799
 DB 711 -----SVETGVAKSITLEDKTKMTKGLVGIYEDKA-----NPEITSEMKKAVSKG 757
 QY 800 LDILITLVEKGPPLAQ-----TNVQGV--YLKTPPLPIPEYVIGLVNYPFKSG 846
 DB 758 VDNSTALPEDKQALKQASAPALDRATQNTFSGIKGONLEPPKRDYI-----NKAQ 810
 QY 847 KLIYALDMSPTIGHQKDAYGNPIIANDVDENEGHTLAVATLADYEGLYIKDILNLSIDK 906
 DB 811 DIAYALK-----NVTTVYLDANPEKRESEBEEVNMKTSIINDISKIAIEK 856
 QY 907 IKAIROIPLAKYHRLGIFQAIRNAAAPADRLPKTPGYINSEVPNVRKKQ--MEKYLKVPD 965
 DB 857 VNNIRAN-LSPDSNL--KTLIEKKAATKVVVELVEFGTKSSTEQSQSPITQNL--ID 910
 QY 966 YKT-----PIFKALPNEKVDGDBRAKGNINAV-----TNSVAVTPI-RSQQLHKS 1013
 DB 911 DKTLSEKVRLOITDKLQEOQKRAEAIKPNVATIEDLRVYSGSALKPISNDEPDIKT 970
 QY 1014 QSDVNLPTQSSKNNFIYEILG 1034
 DB 971 KMVVGRRGRVNIKNI--KIMG 989

RESULT 8
 RBP1_PLAVB STANDARD; PRT; 2869 AA.
 ID RBP1_PLAVB
 AC 000798;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Reticulocyte binding protein 1 precursor.
 GN RBP1.
 OS Plasmodium vivax (strain Belen).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OK NCBI_taxid=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731;
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 mezozoites";
 RL Cell 69:1213-1226 (1992).
 CC -1- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to

CC human reticulocyte cells.
 CC - SUBUNIT: Homodimer (potential).
 CC - SUBCELLULAR LOCATION: Membrane-bound.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).

DR EMBL: M88097; AAA29743.1; -
 KM Malaria; Receptor; Signal; Transmembrane.
 FT SIGNAL 1 17
 FT CHAIN 18 2869
 FT DOMAIN 18 2807
 FT TRANSMEM 2808 2826
 FT DOMAIN 2827 2869
 FT SITE 1030 1032
 FT SITE 2599 2601
 SQ SEQUENCE 2869 AA; 350213 MW; B9DBA42205BECFF CRC64;

Query Match 3.8%; Score 202.5; DB 1; Length 2869;
 Best Local Similarity 17.9%; Pred. No. 0.12;
 Matches 211; Conservative 178; Mismatches 406; Indels 385; Gaps 50;

QY 4 KHLKTLALALT -VSVTYVQGVYGLERESVKOEQTQASDDMF-----EED----- 51
 DB 827 KKKEFSDAFSTKEALQNSMOQNOEGDAIEKKQNRSEKEBEYFNESVEBDSREET 886
 QY 52 -----NEKKNVSKENSTVDVETVSDLFSDNSNNSSKTESVSDPKOVKA- 98
 DB 887 EEOGYTHKKNFRRKEISAEITNMREVINIKIESQUNYGVIEKFSLIGDQNEVSTAK 946
 QY 99 -KPEVTOEASNSNDASKYEVPRKDTASKKETLET--STWEAKDFVTR-----G 144
 DB 947 ALKEKIVSDLRKIDQYETEF--KEKTSAYENTVSTIQSLKSIDSLKRLNGSINNCKY 1005
 QY 145 DTVLGFSGKSGINKLSQTSVHLVPSHAADGQLQVASFATPDCKTAIAEYTSRLGNGK 204
 DB 1006 NTDIDLRSKIKTLREEVOKEMPRKGCENTALLKSLRDMGKINKLN-----DGR 1061
 QY 205 PSRLDIOKHEIDEGELFNAYQLKLIPIGYSIGDAF--VD-----NKNTAEVLP 257
 DB 1062 LNSLDYTKKEDLTK---FYSESKSKTILSKDQK--PDDPLNRIDEMEDIRKDVDELNV-- 1114
 QY 258 SLETTISDYAPAFMSLKQVKLPDNLKVIQELAFPDNOIGKLYLFRHLIKLAERAFKSNRI 317
 DB 1115 -----NYQVISE-----NKV-----TLFKNSV 1132
 QY 318 QYVAFLOSCLKVIGESAFQDNNLNANVLPDGLKIESEAFQNGDEHYNNQVTLRTRTG 377
 DB 1133 TYLEAMSHINTV-----AHGITSN-----KNEILKSYKEV 1163
 QY 378 QNPHQLATENTYVNPDKSLWRATPDMDYTKLSEDFYQKKSATVFGSKGLQKVRKKNL 437
 DB 1164 EDKLNLYEON-----EDYKKNVNP--NEKQLEAIRGMS- 1196
 QY 438 EIPKQNGITITTEIGDANFRNVDFQSKTLRKYLDEIKLPSTIRKIGAFQSNNLKSP 497
 DB 1197 -----KLKEV-----INKHVEEMQLJESTANTL-----KSNAGKE 1227
 QY 498 ASELEKIEGAFNNRIGTL-DLKDTLIKIGDAFIHNIYALVLESVEIGERSAFQ 556
 DB 1228 NEHLEER-----LNKTKGQRDIYEKIKIAE-----LKEGTVELKDNNEK 1270
 QY 557 NGALHMFIGNKVTIGEMAFISNKLSEVNLSEQOKLTIR-----VOAFSDN 604
 DB 1271 -----ANKVPEPRERNITGVLEITVEKQKAGVVEEMMSIKTIKLIQIFSD 1321
 QY 605 ALSEVY--LPPIQTR--EBAFRN-----HLKE-VKGSSTLSQITFNAFDQNDGDKR 653

DB 1322 SQNELVTSITKHLNKAQVEDYIKNREDSIQLEKAKSLTLEDEM----- 1368
 QY 654 FGKVVYVRTHNNSHMLADGRFLIDPKLSTMTVDLKLKLTIEGLDYSL-----RQT 707
 DB 1369 --KGLVQVVMNIQSAIQNAGI-----SKEINELKGVIELLISNYSSTILEYKNS 1420
 QY 708 TQGFPEMT-----TAKALLSKSNLQGEKQKLOEAFGLGRVDDKAIKAKKALVTK 763
 DB 1421 ESVRFQLANGETTAKGEERKASARLAERKLEQIVKQVLDYSDIDDVKYKLE----- 1474
 QY 764 KATNGHLERSINKAVLYNNSAI-----KCANVRLKELDLTLVLR 808
 DB 1475 -----GKRRIILKKKSALTFWESEKFKQKQSSHMENAKGKKIERYLN--N 1521
 QY 809 GKQPLAQATNVQ-----GVYILKTPPLPEYIYGLNYPFKS-----GKLIYALDMSTIG 859
 DB 1522 GDGKKNITDSQEEVGNVYKSA-----EHAFFHVAQVDKTAFCESIVAVYTKMDLFN 1577
 QY 860 EGQKDAVGNPILNVDDNKGYPHTLAVATLADYEG----- 893
 DB 1578 ESLMEKVKYKCEKNDKAEKYS-----AKLKPVGRIKAVSENERKISELKAKYKKE 1633
 QY 894 -IYKDIILNSLDKIKAIRQ-----IPLAKYHRLGIFQAIRNAAAEARLLPTPKG 944
 DB 1634 SSQANDVSTKSLQINCRQQLDSVLSNIGRVQNALQYED-----SADKSKSVLPISLIG 1690
 QY 945 YLNEVENYR--KKQKKNLKPVDYKTPPIPKALPNKQVODRAAKGNINAE-----TNS 998
 DB 1691 AEKSLDKVKAASKSYKKNLETVQ-----NEM-----SRINVEGSLTDID 1730
 QY 999 VAVTPIRSEQQLKHSQSDVNLPG-----TSSKNFIYIILG 1034
 DB 1731 KKTIDLENDLKKKQYBGLQKIKENADKRSNBEIVG 1770

RESULT 9
 SCA4_RICPR STANDARD; PRT; 1022 AA.
 ID SCA4_RICPR Q9ZD49; Q9ZD48;
 AC Q9ZD49; Q9A136; Q9ZD48;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antigenic heat-stable 120 kDa protein (P5120) (120 kDa antigen)
 DE (Protein PS 120).
 GN SCA4 OR RP498/RP499.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OC NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid B.
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.B., Zomoriidipour A., Andersson J.O.,
 RA Sicheitz-Ponten T., Almark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Minkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria";
 RL Nature 396:133-140(1998).
 RL [2]
 RP SEQUENCE OF 11-1016 FROM N.A.
 RA Sekegova Z., Roux V., Raoult D.;
 RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
 RT 'gene D' coding for an intracytoplasmic protein";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -1 SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1 CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 CC FRAMESHIFT IN POSITION 234.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AJ235272; CA14951.1; ALT_FRAME.
 DR EMBL; AJ235272; CA14950.1; ALT_FRAME.
 DR EMBL; AF200340; AAK31305.1; -.
 KW Antigen; Complete proteome.
 FT CONFLICT 11 EFDP L -> RGLV (IN REF. 2).
 FT CONFLICT 365 H -> Y (IN REF. 2).
 FT CONFLICT 413 MISSING (IN REF. 2).
 FT CONFLICT 957 G -> R (IN REF. 2).
 SQ SEQUENCE 1022 AA, 114410 MW, 03230E3A663A9622 CRC64;

Query Match 3.8%; Score 202; DB 1; Length 1022;
 Best Local Similarity 18.8%; Pred. No. 0.031;
 Matches 207; Conservative 196; Mismatches 400; Indels 298; Gaps 54;

QY 28 LERESVQK--QTQASSEDWFEEDNEKTNVSKNSSTVDVSDLPFGDGNSSSK 84
 Db 85 LKQKRDILREYFVNTPPELAEQIAKEEDRKFRFLSNQDNYALINKAFEDTKKMLEK 144
 QY 85 TE-----SVSDPKQVPRAPKEVTOASNSNDASKVEVPKQDTASKKELTETSTWE 136
 Db 145 AEIVGYKNVLSTYSVANGYGGFQVQWENOVASADL-RSTVYKNDSEBELCTINETTVK 203
 QY 137 AKDFVT-RGDTLVGFSK-----SGINKLSQTSHTLVPSHADGTOLOVASFAFPD 187
 Db 204 TKDLIVAKQDGTQVQINSYREINFPKIDKANGSWHLSMVALKADGTAPADKAVYFT-- 261
 QY 188 KKTAAEYTSRLGNGKPSRLDIDQKEIIDSEIFNAVQLTKLTIPNKYKSGIGDAFVDN 247
 Db 262 ---AHYBE--GPNKRP----- 272
 QY 248 KNIAEVNPESETTISDYAFAMSLKQVLPDNLKIVG---ELAFPDNIGKLY--L 300
 Db 273 ---QKEISSPQLKVFVGIGDAVYIHH--GGEITYLAV 307
 QY 301 PPHLTKLARPAKSNRIQTV--EPLGSKLKYTGASFOQDNNLRVAMLPDLEKIESPAF 357
 Db 308 TRGKYKEMWEVALNHGQVALSQTLADLTIVQSPSHETH--KEIIIPN--QELSSIT- 362
 QY 356 TGNPDGHNQV-----VLKRTGONPHQLATE-----NTYVNPDKSLMRATP 401
 Db 363 ---EQTSGQVPTTTNKSLOPKISQ-THQLPQQAQSSGTPNVLNANALSTSMQ 416
 QY 402 DW--DYTKWLEEDFTYQKNSVTGFSNKGLOKVRNNKMLEI-PKONGITTEIGNAPFN 458
 Db 417 DLMNINSYLTNNQDINKQS--DLIKEAIAIILNNKSPFAEKQYV--IIDLAKNIEFN 471
 QY 459 VDPQSKTLKYLEEIKLBSTRKIGAFQSGNNLKSFPASDLEIKGAMNNRIG-- 516
 Db 472 KQIIDA-----KNNVNTL-----LETTONDNTLIDIKSKILEDTVAIT 512
 QY 517 ---TDLKDK--LIKIGDAAPHI-NHIVAYLPESVOEIGRSAPFONGALHMEIGNK 568
 Db 513 LMSNETLKKOKQIILEKVDIGISTODISRVAVDSIMD---TVIKSN-----IABE 562
 QY 569 VKTIGEMAFLS-----NKLESVNLSEOKU-----KTIIVQAFSDNALSEVLPNLQTI 618
 Db 563 DK---EKIFITVDPQINSYEFSVAKOKLDSILKTAETQVLSPE--OQOLANNQMLDNI 617
 QY 619 RBAFAFRNHLKEVKG-----SSTLSQITENAFDQDQDKRPFKKVYVTHNNSHMLADG 672
 Db 618 TTEHTKRDTEKNNIILEPLSNTAKT-----NQVNTSN----- 654
 QY 673 ERFIIPDKLSTMDLEKVLKIEGLDYSTLRQTOFOFREWTIGKALLSKSNLQGE 732
 Db 655 ---VLD-----SPQIEMSKSLIQVVT-KTVAESLVEPKDTEIVKG-IGKTIYTHSD 703
 QY 733 KQKFLQEAQFELGRVDLDKAIAPAEKALVTYKATKNG--HLIERSINKAVLANNALK 789
 Db 704 TSLPLHDKXVIMGSV--AKGIVESKNDLDRRELIIGLVGDIYEAGDNVAVVAISMIA 761

QY 790 KANVKELEK-LDLTLDIYEGKGPLAQAATWQGVYLKTPLPPEYYIIGLVYEDXSGKL 848
 Db 762 NSNINSEKALKRSDQVSEK-----VLDKEIQ-----NIRELKAQN 800
 QY 849 IYALMSDRTIGEGQK---AYGNPILNVEDNEGYHTLAVATLADYGLYKIDILNSSL 905
 Db 801 INSKLHDLIYNKTQVYANLAKNVITTVLDDNSGQREVSEAPKQVSSL-LINDISKRTIE 859
 QY 906 KIRAIROIPLAKYHRLGIFQAIINMAAEADRLPKTPKYLN-----EVENYKQKQENX 960
 Db 860 KINMLRAM-LSDQGNLTFPEEKD---EATKYDELTKVADNKSSTEQQNFIKSNL--- 912
 QY 961 LKPDYKTPILNKALPHEKVDG-----DRAAKGHVIMATNNSVAVT-----PI-RS 1006
 Db 913 ---IDNKT--LSRETRIQIITDNLKQAQKRAETIEMLSAKTEDPRAVSKSELPKISQ 967
 QY 1007 EQOLHKSQSDVNLQOTSSKN 1027
 Db 968 EPIYQAKAVYERNDVIDKN 988

RESULT 10
 ID TE95 TETTH STANDARD; PRT; 872 AA.
 AC 094819;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Telomerase component p95 (EC 2.7.7.-).
 OS Tetrahymena thermophila.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 CC Tetrahymena; Tetrahymena.
 OK NCBI_TaxID=5911;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=95292335; PubMed=7774009;
 RA Collins K., Kobayashi R., Greider C.W.;
 RT "Purification of Tetrahymena telomerase and cloning of genes encoding
 the two protein components of the enzyme."
 RL Cell 81:677-686 (1995).
 CC -1- FUNCTION: Ribonucleoprotein DNA polymerase that catalyzes the de
 novo synthesis of telomeric simple sequence repeats. P95 contains
 some or all of the template-independent primer DNA-binding site
 termed the anchor site.
 CC -1- SUBUNIT: Telomerase consist of two subunit, p80 and p95 that form
 a 1:1:1 complex with the 159 nt telomerase RNA.
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; U25642; AAC66602.1; -.
 DR PIR; S55940; S55940.
 KW Telomerase; RNA-directed DNA polymerase; Telomere; Nuclear protein;
 KW DNA-binding.
 SQ SEQUENCE 872 AA; 102946 MW; 78AF4CD829651FA CRC64;

Query Match 3.7%; Score 200; DB 1; Length 872;
 Best Local Similarity 20.3%; Pred. No. 0.031;
 Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

QY 14 TTVSVVTVYQSEVYGE-----RESVYK-----QEQTQASSEDWFEEDNEK- 55
 Db 17 TMDLFLQNLLEYKKSQIEHYKTOQOOQIIEEDLKLKFNQDQDQNSGNDDEENNSNQ 76
 QY 56 -----TVNSKNSVTVDVSDLPFGDGNSSSKTSVSDPQVPRAPKEVTOASNS 108

Db 77 QELLRRVNOIKQOVOLIKVGSVKYKDLNIN-----EDEN 111
 Qy 109 SSNDASKEVEPKODTASKEETLETSTWEA---KDFVTRGDTLVGFSKSG-----154
 Db 112 KKNGLSSQOV-----KEEOLRTITEOVYQVLMVNMVYQDLNBSGGHRRHRETDY 164
 Qy 155 -INKLSQTSVLVPSHADGTQVLQVASFAPTPDKKALAEYTSRLGDNKPSRLDDQK 213
 Db 165 DTEKWEFISH-----DQKNYSIYA---NOKTSYCMWLDYENKNNYDHLNIN 211
 Qy 214 EIIDEGERIF---NAVQTLKTLIPNGYSIGODAFVNDK-----NIAEV- 253
 Db 212 RLEHAFAPFDDSSQITKLT--NNSYQTVNIDVNFNNLCLALIRFLSLERPNILNIR 270
 Qy 254 -----NLPESLETTSDYAFAMSLKQVYLPDNLKVYIGELAFEDNOIGK-----297
 Db 271 SSYTRNQYNFEKIGELLETIFAUVFSRHILQIHLQVCEAFQYLVNSSQSISVKSQLO 330
 Qy 298 LY-----LPRHLIKLAERAFKSNRIQTVFELGSKLVIGEAS 334
 Db 331 VYSESTDLKLVDTNKKVODYFKELOEPR-LTHVSQKALPVSATNAVENLAVLKKVKA- 388
 Qy 335 FODNNLNWMLPDGLEKIESEAFNGDEHYNNQVLAFTRTGQNPQALATENTYVVPDK 394
 Db 389 -----NMLVSLP-----TQFNFDFPVVNLQHLKLEFGLERPIILTKQKL-----E 428
 Qy 395 SLWRATPDMDYTKLEEDF--TY--QKNSVTGFSKNG--LQKVRNKNLEIPKQHNITIT 449
 Db 429 NLLISTIKQSNLKEFLNLFYTYVAQETSRQIQLKQATITINKNNKQOEESTPEKOTPS 488
 Qy 450 EIGNAFRNVDFOSKTLRKTYDLERIKLPST---IRKGAFAFOSNNLKSEAS--EDLE 504
 Db 489 E-SISGKAFPDHSELTELEDF-SVNLQATQELYDSILKLLINSTLNKFKLSTYKEMEK 546
 Qy 505 IKEGAF--MNNRIGTL-DLXDKLIKIGDAFHNIYAVLPESV-----QEI--G 550
 Db 547 SKMDTFIDLNKYETLNNLRKCSVNISNP--HGHSISELNKOSTFYKFKLTINQELQHA 604
 Qy 551 RSAFRQNGALHMFIGNKYVKTIGMAFLSNKLESV-----NLSEQOKLTTEVOAFS 602
 Db 605 KYTFKON---EFOFNNVAS---AKIESSELSLEBIDSLCKSIASCNNKQOV-----650
 Qy 603 DNALSEVLPNLOTIREEAFKNNHL---KEVGSSTLSQITENA--PDONDGKRREGKY 658
 Db 651 -NIIASILYNNQO---KNEFNKNLFLFKQFOLKULEVNSINCILDO-----695
 Qy 659 VWRTHNSMLADGERPIIDPKLSSTWVLEKVLKIIEGLDYSTLRQTTQ-----709
 Db 696 -----HINSISEFLERKKIKAFILKRYVILQY--LDYTKLFKTLQOLPELNQVY 745
 Qy 710 ---TOPREMTAGKALLSKSLRQGEKX-----FLQBAQFLGRVIDDKALA--KA 756
 Db 746 INOOLEBELVISE---VHKQWENHOKAFYEPICEFIKSSQTLQIDLDQNTVSDSI 801
 Qy 757 EKALVTKATKNGHLLERSINKAVLAVNNSAIKAVYKLEKEIDLTDLVEGKPLAQA 816
 Db 802 KILLESISESKYHIALNPSQ-----SSSLIKSEN-----EIIQELAKRCDKGVILVKA 851
 Qy 817 TMVQGVYLLKTPLPLEPYIYIGLVNVPD 843
 Db 852 -----YKKEPLCLP-----TGTYVD 866
 RESULT 11
 IF2 MOLSU
 ID IF2 MOLSU STANDARD; PRT; 939 AA.
 AC Q7M7X5;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Translation initiation factor IF-2.
 GN INFB OR MS2016.
 OS Wolinella succinogenes.

CC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 CC Helicobacteraceae; Wolinella.
 CC NCBI_Taxid=844;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSMZ 1740;
 RX MEDLINE=22862897; PubMed=14500908;
 RA Baar C., Eppinger M., Raddatz G., Simon U., Lanz C., Klimmek O.,
 RA Nandakumar R., Gross R., Rosinus A., Keller H., Jagtap P., Linke B.,
 RA Meyer F., Lederer H., Schuster S.C.;
 RT "Complete genome sequence and analysis of Wolinella succinogenes";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:11690-11695(2003).
 CC -I- FUNCTION: One of the essential components for the initiation of
 CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
 CC hydrolysis and promotes its binding to the 30S ribosomal subunit.
 CC Also involved in the hydrolysis of GTP during the formation of the
 CC 70S ribosomal complex (by similarity).
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -I- SIMILARITY: Belongs to the IF-2 family.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: BK571662; CA611018.1; -.
 CC HAMAP: MF_001000; -.
 DR PROSITE; PS01176; IF2, 1.
 DR Initiation factor; Protein biosynthesis; GTP-binding;
 KW Complete proteome.
 FT DOMAIN 441 589
 FT NP_BIND 447 454 GTP (BY SIMILARITY).
 FT NP_BIND 493 497 GTP (BY SIMILARITY).
 FT NP_BIND 547 550 GTP (BY SIMILARITY).
 SQ SEQUENCE 939 AA; 102670 MW; B3D7BD9FB5CE3BC0 CRC64;
 Query Match 3.7%; Score 200; DB 1; Length 939;
 Best Local Similarity 19.7%; Pred. No. 0.034;
 Matches 214; Conservative 148; Mismatches 365; Indels 360; Gaps 49;
 Qy 29 ERESVYKQ--EQTQASEDDWFE-EDNERKTVSKENSTVDETVDLPSDGNSSNSKKT 85
 Db 92 EVKESYKAPESLPESPKEAFEAEPKESVKTPTLRLQEPKELVSGISPELSASST 151
 Qy 86 ESVVSDPKQVPAKPEVTOEASNSNDASKVYVPKODTASKEETLETSTWEADQVTRGD 145
 Db 152 ---LSDSNLPQKTP--TKETTVATTLATQTDHEIDSESEKKTLLQATVQK-----200
 Qy 146 TLVGFSGKGINLQTSHTVLVPSHADGTQVLQVASFAPTPDKKALAEYTSRLGNGKP 205
 Db 201 ---VGLRI--YKRSSE-----PAKADRSLEB---ARTPSTAGAKTLOSLGES---244
 Qy 206 SRLLDQKSLIDEGEFLNAYQLTK-----LTPNGYKSGIGODAFVNNKNTAEVNL--PES 258
 Db 245 -----DESEALARKKKKKKKKPLPAPT-----KKKQOKIDLLGDRA 281
 Qy 259 LETISDYAFAMSLKQVLPDNLKVYIGELAFEDNOIGKLYLPRHLIKLAERAFKSNRIQ 318
 Db 282 LETVSSF-----DDEQE---EIVLPULT-----RDLINKEDVAK--KVD 317
 Qy 319 TVEFLGSKLVIGEASFODNNLNWMLPDGLEKIESEAFNGDEHYNNQVLAFTRTGQ 378
 Db 318 T---DRIKVRKTFPLQGIIRV-----KRRKR 342
 Qy 379 NPHQALATENTYVVPDKSLWRATPDMDYTKLEEDFTYQKNSVTGFSKNGKLYKVRKNLE 438
 Db 343 RPEYTA-----DKESISGTEIRP---EIRAYVEPAEKKGKIGEVYKVLFINGLM 399
 Qy 439 IPKQHNITITTEIGNAFRNVDFOSKTLRKTYDLERIKLPSTIRKIG-AFAFOSNNLKSPFE 497

```

Db 390 ITKN-----DFLDRDSIRILAELELDVIVIKNTSEALREYSEB-----E 428
Qy 498 ASDELRIKAGAFMNRIGTLD-----LKDKL-----IKIDDAFHNIHAIYLPBESVQ 548
Db 429 EDEBEELREPRPVYITMGVHDGKTSLLDKIRNTKVAABEAGITQHIAVYVEKDKK 488
Qy 549 IGRSAFQONALHLMFIGNKVKITIGEMAFLSNKLSEVNLSEQOKLTIEVQAFSDNALSE 608
Db 489 I-----SFIDTPGHAFTEMRARGEVND-----IVIVIAADDOVKQ 526
Qy 609 VLPENLOTIREAKFNHLKEVGSSTLSQITFNAPDQNDGKRFKGVVATHNNSHM 668
Db 527 -----QTI--EA-----LNHAKAANVPITIALNKVDKDDA----- 554
Qy 669 LADGERITDDPLDLSSTWVLEKVKLIEGLDVSITLQOTQOTQOREMTAGKALLSKSNL 728
Db 555 -----NPDVKKAAAD-----LGSPLEWGEYEVHIS----- 583
Qy 729 ROGKQKQLOBAQFELGRVDLDKAIARAKALVYKATKNGHLERSINKAVLANNNAI 788
Db 584 -----AKTGEGI-----DHILE-----TILVQSELE 605
Qy 789 KKANVKLEKELDLTDLVBEGKPLAQTWQVYLLKTPLEPYITGLNVYFDKSGKL 848
Db 606 LKANPERKAAVAVIESLIEKKGKGVAVIVQSGTLKVDSDIVADTAGRYVALIDDCGN 665
Qy 849 IYALDMSD-----TIGEGKDAVGNPILVNDENEGHYTLAVTADYGLYI--KDILNS- 902
Db 666 IQSTIGPEVAVVTLSTETPMAGAVLVSEVDS-----IAREVAKRALYLRQKELSRST 719
Qy 903 --SLDKTKA-----IROIP-LAKYHLGIFQAIRNA-----AAE 933
Db 720 KVSFDELISAWABEQGLMSLPITLADQSGLEAIRSGLEKRNBEVKINIHAQVGITTE 779
Qy 934 ADRLPTPKCYL-----NEVYNKCKOMEKYLKPYDYKTPFENKALPNEKVDGKRAKH 989
Db 780 SDVTLAASDMSVILGFRVREPTGSVKRAKEL--GEVYVYSIIVL-----LDDVRAVLGG 834
Qy 990 NINMETNNSVAVPFIRESQOLHKQSVDVNPQTSKKNFTIELIGVSLCLFVTVAG-- 1047
Db 835 -----MMSVLEBE-----NTQAEVREFTTIAKVGTIAGC--LVTDGSI 872
Qy 1048 KKGKRRAR 1054
Db 873 QRGTKVR 879

```

```

RA Renz M.;
RT "Macrogolgin -- a new 376 kD Golgi complex outer membrane protein as
RT target of antibodies in patients with rheumatic diseases and HIV
RT infections.";
RL J. Autoimmun. 7:67-91(1994).
RN (3)
RP SEQUENCE FROM N.A.
RX MEDLINE=95100974; PubMed=7802676;
RA Schda M., Misumi Y., Fujiwara T., Nishioke M., Ikehara Y.;
RT "Molecular cloning and sequence analysis of a human 372-kDa protein
RT localized in the Golgi complex.";
RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).
CC -!- FUNCTION: May participate in forming intercisternal cross-bridges
CC of the Golgi complex.
CC -!- SUBUNIT: Homodimer; disulfide-linked.
CC -!- SUBCELLULAR LOCATION: Golgi; membrane-associated.
CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the
CC autoimmune disease Sjogren's syndrome.
CC -!- SIMILARITY: Belongs to the golgin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, X75304; CA53052.1; -.
DR EMBL, D25542; BAA05025.1; -.
DR PIR, A56539; A56539.
DR PIR, I52300; I52300.
DR Genew; HGNC:4429; GOLGBL.
DR MIM; 602500.
DR
DR GO; GO:0000139; C:Golgi membrane; TAS.
DR GO; GO:0005795; C:Golgi stack; TAS.
DR GO; GO:0016021; C:Integral to membrane; TAS.
DR GO; GO:0007030; P:Golgi organization and biogenesis; TAS.
KM Golgi stack; Antigen; Coiled coil; Transmembrane.
FT DOMAIN 1 3235
FT TRANSMEM 3236 3256
FT DOMAIN 3257 3259
FT TRANS 48 593
FT DOMAIN 677 1028
FT DOMAIN 1062 1245
FT DOMAIN 1301 1779
FT DOMAIN 1828 3185
FT DOMAIN 2420 2423
FT DOMAIN 2996
FT DOMAIN 2996
FT CONFLICT 1 39
FT CONFLICT 215 215 A -> AQLSSM (IN REF. 3).
FT CONFLICT 1765 1765 D -> G (IN REF. 3).
FT CONFLICT 2950 2950 H -> D (IN REF. 3).
SQ SEQUENCE 3259 AA; 376075 MW; 60376420DBA178DD CRC64;

Query Match 3.7%; Score 199.5; DB 1; Length 3259;
Best Local Similarity 20.7%; Pred. No. 0.2;
Matches 217; Conservative 168; Mismatches 386; Indels 277; Gaps 52;

```

Db 464 --AVTEENIASIQKRVEL-ENKGAALLSSIELEBKAKENKLS--QTLLBAQRTEG 519
 QY 235 GYSGIOGDAFVNDKNTAELAVNLPESLESTISDYAPA--HMSIKQYKLDNKLKIGELAFPN 292
 Db 520 ADEVESEISTVDIANRSSSSAESGGQVLENTFSQKHEKEL-SYLLLEMKEAQSEIAFLKL 578
 QY 293 QIQGKLYLPNHLIKLAERA-----FKSNRIQTVEFLG--SKLYTIGASFOQNNLNKVM 345
 Db 579 QLOG-----KRAEADHEVLDDQEKQKQMGEGIAPIKMYVLEDQDPP-----M 625
 QY 346 PD--GLEKIESAFTGPNPDEHYNNQVLRTRTQ--NPHQALTENTYVNDKSLWR 398
 Db 626 PNEESSIPAVEKE-----QASTER-----QSRTESEISLNDAGVELSKTKQGDKSL-S 673
 QY 399 ATDDMDYTKWLEDPTQKRSVTFGSKGLOKTRKANKLEIPQOHGITTETIGDAPRN 458
 Db 674 AVEDIQCH--QDELEBKSKQIL-----ELET-NFHAQEIYE-----KN 710
 QY 459 VDFOSKTLRYD--LEBIKLPTIRKIGAPAFQSNMLKSFEA-SED-----LEBIKEGAPM 511
 Db 711 LDKAKEISLNLQIEEFKNA-----DNSSAFTALSEEDQLSQVKELSWV 759
 QY 512 NN--RIGTLDL-----KDKLIKIGDAFHINIYAIPLPESVQETIGSAFONGALHL 562
 Db 760 TELRAQVQKLEMMIAEAEERQRLDYESQTAHDN-----LITTEQI-----HS 800
 QY 563 MFIGNKYKTIGEMAFLSNKLKESVNLSEQKQKTI-EVQAQSDNALSRVLPVLTITREE 621
 Db 801 LSTIAKSKDV-KLEVQNELDVLQOLPSEQSTLIRSLQOLQNKESVLT-----EG 850
 QY 622 AFRKRNHLKEVKGSSSTLSQITFNAFPQNDGDKRFKQVYVTHNNSHMLADGERFIIDPDK 681
 Db 851 AERVRH-----SSKYEEL-----SQALSGKELIETMDQ 880
 QY 682 LSTWMDLEKVLKTIISGLDYSTRQTTQTOFREMTAGKALSKSNLROGKCKPIQEAQ 741
 Db 881 IL-----LEK-----KRDVETLQOTIEEDQCVT-----EISFSMEKQVQINEEK 921
 QY 742 FPLGRVLDLKAIAKBAKALYTKATKNGHLEERSINKAVLANNSAIAKKA---NVRLEK 798
 Db 922 FSLG-VET-KTLSEQLNLIRABEAKKQVVEDNEVSSGKONYDEMSPAGQISKELOH 979
 QY 799 ELDLTLVYEGKPLAQATVWQVYLLKTPPLPEYVYIGLVYFDSKGLIYALDMSDT 857
 Db 980 EFDLLKKNENQRKQALINRKEILQVRSLBEELANK---DESKK---ELIPSETE 1033
 QY 858 IGGQGDAYGNPILNDEDENGYHTLAVALTADYEGLYTQIILNSSIDKTKAIRQIPLAK 917
 Db 1034 RGEVEE-----DKENKEVSEKCVTSKCEIEIYLQITSEKVEVLOHTRKOLEEK 1083
 QY 918 YHRLGTFQATRNAAEADRLPTPKGYLNEVPNRYKKQMKULKPDVYTPIFENKALPN 977
 Db 1084 L-----AAEFOFQALVK---QNMQTLQDKTNQDILQAIISENQAIIOKLITS 1128
 QY 978 --EKVDDRAKAGHINIAETNNSVAVTP 1003
 Db 1129 NTDAQDGSVAL-----VKETVVISP 1149

RESULT 13
 KNTN1_HUMAN
 ID KNTN1_HUMAN STANDARD; PRT: 1357 AA.
 AC 086UP2; Q14707; Q15387; Q86W57;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE Kinectin (Kinesin receptor) (CG-1 antigen).
 GN KNTN1 OR CG1 OR KIAA0004.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Lymphoid;
 RX MEDLINE=95306853; PubMed=7787243;
 RA Fueterer A., Kruppa G., Kraemer B., Lemke H., Kroenke M.;
 RT "Molecular cloning and characterization of human kinectin.";
 RL Mol. Biol. Cell 6:161-170(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Peripheral blood lymphocytes;
 RX MEDLINE=94314220; PubMed=8039706;
 RA Print C.G., Leung E., Harrison J.E.B., Watson J.D., Krissansen G.W.;
 RT "Cloning of a gene encoding a human leukocyte protein characterised by
 RT extensive heptad repeats.";
 RL Gene 144:221-228(1994).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Wang H.-C., Chen W.-F., Su Y.-R.;
 RT "Identification of a variant of Homo sapiens kinectin mRNA";
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Bone marrow;
 RX MEDLINE=96051387; PubMed=7584026;
 RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
 RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
 RT "Prediction of the coding sequences of unidentified human genes. I.
 RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
 RT analysis of randomly sampled cDNA clones from human immature myeloid
 RT cell line KG-1.";
 RL DNA Res. 1:27-35(1994).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX PubMed=12508121;
 RA Heilig R., Beckenber R., Petit J.-L., Fonknechten N., Da Silva C.,
 RA Catalico L., Levy M., Barde V., De Bernardis V., Ureca-Vidal A.,
 RA Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,
 RA Sun H., Du H., Pepin K., Artiguenave F., Robert C., Crouad C.,
 RA Bruneis T., Jailion O., Friedlander L., Samson G., Broctier P.,
 RA Cure S., Seguren B., Aniere F., Samain S., Crespeau H., Abbaei N.,
 RA Alich N., Bocus D., Dichtoff R., Dore M., Dubois I., Friedman C.,
 RA Gouyenoux M., James R., Madan A., Maitrey-Bertrada B., Mangenot S.,
 RA Martins N., Menard M., Oztas S., Ratcliffe A., Shaffer T., Trask B.,
 RA Vachier B., Bellemere C., Belser C., Besnard-Gonnet M.,
 RA Bartol-Mavel D., Boudard M., Briet-Silla S., Combette S.,
 RA Dufosse-Laurent V., Perron C., Lechaplais C., Louesse C., Musielat D.,
 RA Magdeleat G., Patrau E., Petit E., Sivain-Trukiewicz P., Trybou A.,
 RA Vega-Czaray N., Bataille E., Bluet E., Bortelais I., Dubois M.,
 RA Dumont C., Guerin T., Hatray S., Hamadi R., Munga U., Pellouin V.,
 RA Robert D., Wunderle E., Gauguet G., Roy A., Sainte-Marthe L.,
 RA Verdier J., Verdier-Discala C., Hillier L.W., Fulton L., McPherson J.,
 RA Matsuda F., Wilson R., Scarpelli C., Gnapay G., Wincker P., Saurin W.,
 RA Queller F., Waterston R., Hood L., Weissbach J.,
 RT "The DNA sequence and analysis of human chromosome 14.";
 RL Nature 421:601-607(2003).
 RN [6]
 RP SEQUENCE OF 1-870 FROM N.A. (ISOFORMS 1/2).
 RC TISSUE=Lung;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Maruska K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Yoshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., Mcwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
 RA Villalón D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallius D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.",
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [7]
 RP SEQUENCE OF 191-195; 395-405; 545-554; 747-761 AND 811-823.
 RA MEDLINE=21969647; PubMed=11973345.
 RA Tian H., Fankov R., Tian S.D., Hampton B., Burgess W.H., Yamada K.M.,
 RT "Integrin clustering induces Kinectin accumulation.",
 RL J. Cell Sci. 115:2031-2040(2002).
 RN [8]
 RP CHROMOSOMAL LOCATION.
 RA MEDLINE=96163023; PubMed=8575822;
 RA Print C.G., Morris C.M., Spurr N.K., Rooke L., Kristansen G.W.,
 RT "The CG-1 gene, a member of the Kinectin and ES/130 family, maps to
 RT human chromosome band 14q22.",
 RL Immunogenetics 43:227-229(1996).
 CC -1- FUNCTION: Receptor for kinesin thus involved in kinesin-driven
 CC vesicle motility. Accumulates in integrin-based adhesion complexes
 CC (IAC) upon integrin aggregation by fibronectin.
 CC -1- SUBUNIT: Parallel homodimers formed between the membrane-bound and
 CC the cytosolic form, and also between 2 cytosolic forms (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Vesicle membrane
 CC protein anchored to the endoplasmic reticulum.
 CC -1- ALTERNATIVE PRODUCTS: Named isoforms=2;
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q86UP2-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q86UP2-2; Sequence=VSP_007981, VSP_007982;
 CC -1- TISSUE SPECIFICITY: High levels in peripheral blood lymphocytes,
 CC testis and ovary, lower levels in spleen, thymus, prostate, small
 CC intestine and colon.
 CC -1- SIMILARITY: Belongs to the Kinectin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

FT CARBOHYD 1055 1055 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1068 1068 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1263 1263 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1329 1329 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPPLIC 1031 1059 Missing (in isoform 2).
 FT VARSPPLIC 1232 1259 /FTId=VSP_007981.
 FT VARSPPLIC 282 282 Missing (in isoform 2).
 FT VARIANT 282 282 /FTId=VSP_007982.
 FT VARIANT 282 282 V -> M (in dBSNP:2274073).
 FT CONFLICT 15 15 S -> P (in REF. 4).
 FT CONFLICT 210 210 MISSING (in REF. 1).
 FT CONFLICT 373 373 I -> M (in REF. 1).
 FT CONFLICT 939 939 E -> G (in REF. 1).
 SQ SEQUENCE 1357 AA; 156274 MW; 971FCDF8A8FC88E CRC64;
 Query Match 3.7%; Score 199; DB 1; Length 1357;
 Best Local Similarity 19.2%; Pred. No. 0.062;
 Matches 218; Conservative 210; Mismatches 404; Indels 302; Gaps 59;
 2 TKKHLKTLALATLTVS-----VVTYSQEVYG-LEERSVSKQEQTSASDDW-----FEE 50
 351 TKORCKQLTQPMTEKERSNVVTRMKORIGTLKEHNVQNKLIHVSQETQOMKFPQ 410
 51 DNEKK-----TWSKENVSTVEDVSDLPDSGNSNSSSKTESVSDPPQVPAKKEVTQEA 106
 411 VREQMEALVHLKQENGILDAVSN-----TTNLESKQSAELNKRQ---DYARLVNEL 462
 107 SNSNSDSKVEVPQODPAKSKETLETSTWEKDPVTGTGLTVGSKSGINLSQTSNHLVL 166
 463 TEKGLKQOEVEQCKNEQAATQKVLQOEAE----- 494
 167 PSHADGTQLQVASFATPDKTAIAEYTSRLGENGKPSRLIDQKEIIDEGEIFNAYQ 226
 495 -----RRMEVSYI--RKRALHEH-----AAQDQSKFVAKENVEQSHH- 533
 227 LTKLTPNGYKSIQDAFVDNKNIAE-----VNPESLETISDYAPAHMSLKQ 274
 534 SKLT-----DLTVSKQLEQRIMQMESEQRVKNESLQ-----MGOVD 573
 275 VKLDNKLKVGELAFPNQIGKK--YLPHLILKLAERAKSRRI-QTVEFLSKLKV 330
 574 I-LQNEMLKXQIQOFSQIAQFSASVLAELHKVI--AKKDQIKQTEBSLSERDL 630
 331 --GEASFQDNNLRVWMLPDGLEKTESEAFNPGEDEHYNNQVLRITGTQNPQHV--ATE 386
 631 TSKEBELKDIONMNFILKAVQKIQALA-----NEGAAN-----AHELEKMQQ 673
 387 NTYYNPDKSLWRATPDMQTYWLEBEDFTYQ-KNSVTG---SNKGLQKVRNKNLEIP 440
 674 SVYYVDDK-----IRLLEQQLQHEISNKKWEKFIINDQKALKSEVQKQLTVLS 722
 441 KQNHGITTTEIGDAFRVDFOSKTIKRYDLE--IKLPSTIRKIGAPAFQSNMLKSPEA 498
 723 EOPNPDVEQW-EKCIQEKDEKLTVBEL-LETGLIQVATEEBELNARTESSI----- 775
 499 SEDLEIEKGFPMNNRITGLDILKXLIKTI---GPAAFINRIYALVPEVQOEIGSAR 555
 776 TKHYVDLK--AKQNDQVSFAFLVEHLKVHKKQKI-----KSYVELLEA--- 819
 556 ONGALHLMFIGNKVTI---GEAFLSNKLESVNLSSQKQK-TIEVQASDNLALSEV 610
 820 -----ELIKVANKETVODLKOELVLAEEIGNVQLEKRAQOISLISKVOELQ----- 866
 611 LPPNQLTIREAEFRKNHLEKVE-----KGSSTLSQTFNAPFQNDGKFRGKKVVRTHNN 665
 867 ---NLKKEKE--QNNIKAVLEEKQKLANTKGLQDLOGBENESLKAHVQV---AQHN 918
 666 SHMLADGERFTIPDKLSITWVDEKVLKTIIGL-----DYSTLRQTTQCFREMTTAA 719
 919 LKASASQF-----BELEIVLKEKENELKRLKAMKERESEDSLSKTLQIQ---DVODEN 970
 720 KALLSKSNLRQGEKQKFLQEAQFLGRVDDLKAIKAKKAL-----VTKKAT 766

Db 971 K--LFKQIOLKQONV--QOASFPPEHLLKVISEKEKISGLMELSLKQAVENGRK 1027
 Qy 767 KNGHLERSINKAVLANNNAIKKANKRLEKEIDLTIDVEGKPILOQTMQGVYLLK 826
 Db 1028 KNDLREKKNEMEAALASTERMLODKNKTSKEROOOVEAVE-----LEAKEVLLKLPK 1082
 Qy 827 TPLP-----LPEVYIGLVNVPKSGKLYALDMSPDTTIGEGKADVGNPILNWDENEGYHT 882
 Db 1083 VSPVSNLSYGEWMLG-----FEKRAKEC-----MACTSSSEVKYLEKLEADE---MHT 1130
 Qy 883 LA-----VATLADYGLVYIKDILNLSLCKIKAIROIPLAKTRIL--GIFQAI RNAAAE 933
 Db 1131 LLECECKYKSVLAETEGIILOK--LQSVGEQENKMKVKVDESHKTIKQOVSPTSSEGE 1188
 Qy 934 ADRLPTPKGYLVNVPYKVKQ--MEKMLKPDV-----YKTP-----FNKRLPN 977
 Db 1189 LERLRSEN-----KDLNLRERREHLEMELEKEMERSTVYTVRELKDLTLTELQKLD 1243
 Qy 978 EKYDGRPAKGNH--INAEITNNSVAVTPIRSEQ-----QLHKSQDVNL.PQT 1022
 Db 1244 SYSRAVQNEIMNLKQOLNET--LTQLRTQNERQVADILHKAQOSLILIGS 1295

RESULT 14
 SPOF SCHPO STANDARD; PRT; 1957 AA.
 AC 010411; 09USE9;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Spoolation-specific protein 15.
 GN SPO15 OR SPAC1F3.06C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Bakayota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 ON NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
 RX MEDLINE=20107136; PubMed=10639340; Shimoda C.;
 RA Ikemoto S., Nakamura T., Kubo M., Shimoda C.;
 RT "S. pombe spoolation-specific coiled-coil protein SPO15 is localized
 to the spindle pole body and essential for its modification."; J.
 RL Cell Sci. 113:545-554(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brookes K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Delwail T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Shelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymopiez B.,
 RA Welfens I., Vliegstra E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Filtz C., Holzer B., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Beier P., Zimmermann W., Medler H., Wambutt K., Furnelle B.,
 RA Goffeart F., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Garniti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 RN [3]
 RP SEQUENCE OF 705-871 FROM N.A.
 RC STRAIN=968 h90;
 RX MEDLINE=20223868; PubMed=10759889;
 RA Ding D.-Q., Tomita Y., Yamamoto A., Chikahide Y., Harguchi T.,
 RA Hirakawa Y.;
 RT "Large-scale screening of intracellular protein localization in living
 RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";
 RL Gene Cells 5:169-190(2000).
 CC - FUNCTION: Has a role in the initiation of spore membrane
 CC formation.
 CC - SUBUNIT: Monomer.
 CC - SUBCELLULAR LOCATION: Spindle pole body.
 CC - SIMILARITY: Belongs to the Mpc70 family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z70690; CA94624.1; -.
 CC DR EMBL: AB027811; BAA87115.1; -.
 CC DR PIR: T38077; T38077.
 CC DR GeneDB: Spombe; SPAC1F3.06c; -.
 CC KW Spoolation; Coiled coil.
 CC FT DOMAIN 199 785
 CC FT DOMAIN 804 1235 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1320 1471 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1481 1722 COILED COIL (POTENTIAL).
 CC SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;
 CC
 CC Query Match 3.7%; Score 199; DB 1; Length 1957;
 CC Best Local Similarity 19.4%; Pred. No. 0.1;
 CC Matches 227; Conservative 204; Mismatches 449; Indels 292; Gaps 54;
 Qy 2 TKHLKTLALATTVSVVTVYSGEVGLERBSYKQOTQASDDMFEDNEKRTVSKS 61
 Db 575 SKNSLQTLNAPFOEKLAKSMQ-----LKENQNFSSIDTFFKLNESHQELNN 624
 Qy 62 NSTVDVTSGLDFSDGNSNSSTESVSDPKVPYAKPEVYQEAANSNDAAKAVPKQ 121
 Db 625 HQTITQOLK-----TSSKLOOL-----OLERANFEQKSTLSDENNDLRTKLKL 670
 Qy 122 DTAKK-----ETLETSTWEAKDFVTRGDTLVGFSKGINKLQTSHTLVPSHAADG 173
 Db 671 EESNKSLIKQEDVDLKNKIQLKEDIRSEBALFKLEAKNLAEVIDNLKGRHETLE 730
 Qy 174 TQLTQVASFAPFYDDKTAIAEYTSRLGENGKP--SLDIDOKIILIDGE-----IFNA 224
 Db 731 AQNDNLHS--SLDAKNTVAILSELTSSSEVDKRLTANVETLTQDSKMKQSFSLVNS 788
 Qy 225 YQTLTKLTIPGYSGIQD-AFVQNKIAEVLNPESETTSDVAFAMWSLKQVLPNKL- 282
 Db 789 YQ-----SISNLYHELDHVNMQSNNVTLLSSSKLTIDCE-----NLTOQNTLIDNVQK 940
 Qy 283 -----VIGELAFPDQIGKLYLPRLIKLAERAFKSNRIQVTEFLGSKLVIGE 332
 Db 841 IMEKHVNOGSKESEL-----KEVNGKSLIDKLNKSSLVANVAILSDNDQILTLQABLKS----- 892
 Qy 333 ASFQDNKLNKVMPLPDGLKTESE-----AFTNPGDEHYNNQVAVLRTYT 376
 Db 893 -NVDSDQESQAQNSGLKSLAEKQDLHTENEELHRLDGLTKELTEESKSDLGKRLT 951
 Qy 377 GQNFH--QLATEVTYVNPDKSLWRATPDMDYT-----KWLIEDPTYOQNSVYFGSNKGLQK 430
 Db 952 ARGEHLSNLKEEN--MSQCATISVYSKADETLSKSKLEADLHLKKNVSEVVERNAL 1009
 Qy 431 VRNNKMLEIPKQNHGITI-----TEIGDNAPFRNVDPQSK--TKRYDLEELIKPLPTIRKIGA 485

```

Db      1010 LANSERLMDLKNNGENIASLQTELEKKAENDLQSKLSVSSEYNNLLISS-QTNKS 1068
Qy      486 PAFQSNLIKSEFE-----ASEDLBEI-----KEGAFNNRIGTL----- 518
Db      1069 LEDTQNLQTKLEKVVQKLEKQDQNVNELTSKYGKLGSENAQIDELTALRKSKQ 1128
Qy      519 -----DLKDLITIGGAFFHINHTVALPESQGEIGRSFRONGALHMFIGNKV 569
Db      1129 HDLCANFVDLKEK-----SDLEQ-----LTNEKNEILVLSQSN-----SNNE 1168
Qy      570 KITGEMAFSLNKLSEVNIASEQKQKLTIEVOAFSNALSEVLPENLOTIREAKFNHLK 629
Db      1169 ALVBERSDLANRLSDM-----KKSLS-----SNVNS-VIRSDVAVADEL-----DTLK 1213
Qy      630 EVKGS-----STLSQTFNAPDQNDG-DKRPGRKV-----VRTHNN-----SHMLDGER 674
Db      1214 KDKDLSLSTQSEVQDQDDLDLSLKGESESNKYAVSLREICTSEIDVPESEILD--N 1271
Qy      675 FIIDPDKLSS-----TMDLEKVKLTIEGLDYSTLR-----QTOTQREMTTAKKALSK 725
Db      1272 FVFANQFSLSLRLTVLSLENYLDAFNQVNFKKELDNRLTTTDAFTKVADLEKJ--- 1328
Qy      726 SNLRQGEKQKFLQEAQFFLGRVDIDKAIKAAEKALVTYKA--TKNGHLLR--SSINKAVL 781
Db      1329 -----QHEDHDL-----IQRGDEKALKQSEKFNLEKKAEMTENHISLEGEKETEKEI 1378
Qy      782 AYNSNAIK-----KANVKLEKELDLITDLVEGKGPLAQATMVGAYLLKTPPLP 832
Db      1379 AELSLRLDQNALATNKKLNQDLHNOETRLKEVDLKEK----- 1416
Qy      833 EYVIGLWVYFDKSGKLITVALDMSPTIGSGQKDAVGNLIVNVEDNEGHTTAVATLADYE 892
Db      1417 -----ESLITSLSES-LSNORQKES-----SLIDAKNELEHMLDTSRKNSS 1457
Qy      893 GLYTKDILNSSLDR-----IKAIRQI-PLAKYH--RLGIFQAIRMAAEDRLLPKTPKG 944
Db      1458 IMEKIESINSLSLDKSEFELASAVEKLGALQGLHSESLIMENIKSQLOEAKERI-QVDES 1516
Qy      945 YLNVVP-----NYRKKOMEKN--LKPVDYKPTIPNKALPNKVDGDRAKHNINAE 994
Db      1517 TIQELDHEITASKNNYSGKMDKDSIITRUSENIEQLNLLAEK--SAVK--RLSTE 1570
Qy      995 TNSGVAATPIR-SEQQLHKSQSDVNLPTQSSK 1025
Db      1571 KESRILOENSRLLADLEHFKSQVESELSKSLK 1602

RESULT 15
USOL_YEAST STANDARD: PRT: 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Intracellular protein transport protein USOL.
GN USOL OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota, Fungi, Ascomycota, Saccharomycotina, Saccharomycetes,
OC Saccharomycetales, Saccharomycetaceae, Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260 (1991).
RN [2]
RN SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McCellan M., Tao N.,

```

```

RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Required for protein transport from the ER to the Golgi
CC complex.
CC -1- SUBCELLULAR LOCATION: CYTOSOL; ASSOCIATED WITH INTRACELLULAR
CC MEMBRANS. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, composed
CC of a heptapeptide repeat pattern characteristic of alpha-helical
CC coiled coils. May form filamentous structures in the cell.
CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YDL047C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54378; CA38253.1; -
DR EMBL; L03188; AAB00143.1; -
DR EMBL; U53668; AAB6659.1; -
DR GenBank; 140300; -
DR SCD; S0002216; USOL.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR006955; Usol p115 C.
DR InterPro; IPR006953; Usol p115_head.
DR Pfam; PF04871; Usol_p115_C_1.
DR Pfam; PF04869; Usol_p115_head_1.
DR Transprot; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
KW DOMAIN 1
KW DOMAIN 2
FT DOMAIN 1 725 1790
FT DOMAIN 2 465 1487
FT DOMAIN 1 991 1790
FT DOMAIN 2 1172 1786
FT CONFICT 847 847
FT CONFICT 924 924
FT CONFICT 1253 1253
FT CONFICT 1319 1319
FT CONFICT 1461 1461
FT CONFICT 1581 1581
FT CONFICT 1600 1600
FT CONFICT 1661 1661
FT CONFICT 1772 1772
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B21639FD4818 CRC64;

Query Match 3.7%; Score 198.5; DB 1; Length 1790;
Best Local Similarity 19.3%; Pred. No. 0.096;
Matches 232; Conservative 209; Mismatches 432; Indels 329; Gaps 57;

Qy 16 VSVVYQSEYVGLER-ESVQEQTQASQSDWFEED-NERKTVNSKXENSTVDEVSDF 73
Db 724 ISKISF-EVYBKIQROCTKAKGELTISQTESTHEMLTEKLAIALNHEHLEBKYL- 781
Qy 74 SDGNSNNSKTESV--SDPKQVPAKPEVYQ-----EASNSNSDAKVE---VYKQD 122
Db 782 ---NSSHSILKENFSILETELKVRNDSIDEMTQRLVDLEFYTKENQFALLEYKSTIRKQ 838
Qy 123 TASKK-----ETLETSTWEAKDPYTR-GDILVGS-----KSGINKL 158
Db 839 DSITLKEGLFTTISQKKKADGINKKQDIPALSRMVAEENCKQLQKEKDSNVNHQ 898
Qy 159 SQTSLVLPSPHADQTLQVASFATPDDKTAIAEYT--SRIGENKPSRLDID----- 211
Db 899 KETKSL-----KEDIAAKITETKAIINENBEKKIQCNNSKE 935
Qy 212 ---QKEIIDEGEIFNAVQ-LTKLT-----IPNGYSIGQD-----AFVDNKIAIEV- 253

```

Db 936 KEHISELVEYKSRFOSHDLVAKLTEKLSLANNYKDMQANESLIKAVEESKNSSIQ 995
Qy 254 --NLPSLETTIS-----DYAFAMSLKQVLPDLNKVITGELAFPDNOIGK----- 297
Db 996 ISMLONKIDMSQEKENFOIERGSEIKENIQOLK-----KTSIDLEQTEKEHIIKSDSSKD 1050
Qy 298 ----LYLPHHLIKLAERAFKSNRIQVEFLGSKLVIGF--ASFQDNMLANWLPDGLK 351
Db 1051 EYSSQISLKEKELTATTANDENVNKISELTKREBELAELAAYK--NLKN-ELBTYLET 1107
Qy 352 IESEAFQNGPDGHYNNQVVLRTRTGQNPQOLATENTYVNPDKSLMRATPPMDYTKWLE 411
Db 1108 SEKALEVENEHEHLKEKI-----QLEKAEFTQQLNSLRA--NLESLEKEHE 1155
Qy 412 DFTYQ--KNSVYTGSSNGLOKVRNKNLEIPKQNHGTTITIBIGNAFR---NVDPOSKTL 466
Db 1156 DLAAQLEKYEEOIANK-----EROYNEEISQLNDEITSQOENESIKKNDLEGEVKAM 1210
Qy 467 RKYDLEIRIKLPSTIRKIGAFAPQSNMLKSFEASED---LEIRKGAFFMNNRIGTLDKDK 523
Db 1211 KSTSEOSNLKKS--EIDLNLQIKELKKNETNEASLESIK-----SVESF 1256
Qy 524 LKIGDAAFHINIYALVLESVOEIGRSAPQNGALHLMFIGNVK--TIGEMAFLSNK 581
Db 1257 TVKIKELQDECNFKEKEV--SELEDKLASSEDKNKYLELOKESEKIKELIDAKTTLEKIQ 1315
Qy 582 LESV-NLSEBQOLKIEVOAFS-----DNALSEVYLPNLOTIREAFKRNHLKEVK 633
Db 1316 LEKITNLKAKESESESLKTKTSSEERNAEOLKNEIIOIKQAFEKERKLINEG 1375
Qy 634 SSTLSQITFNAFQNDGDKRFKKV-----VVRTNNSHMLADGERFIIDPKLSSTW 687
Db 1376 SSTITQ-----EYSEKINTLEDELIRLOENENELKA-----KEIDNTRS 1413
Qy 688 DLEKVLKIEGLDYSLRQTTQOFR-----EMTTAGKALLS--KSNLROGEKQ 734
Db 1414 ELEKV-----SLNDELLEKEKQNTIKSLQDEILSYKOKITRNDEKLLSTIERDNKRODESL 1468
Qy 735 K-FLQEAQFLGRVLDKALAKAKAL--VTKATKNGHLLERS-----LNKA 779
Db 1469 KEQIRAAQ-----ESKAKVEEGIKLEESSSEKAELEKSKEMMKLESTIESNET 1519
Qy 780 VLAYNNSAIKKA-----NVRKLEKE-IDLLTDLVEGKPIAQATWQGVYLL 825
Db 1520 ELKSSMETIRKSDKXLEQSKSAEEDIKNLQHEKSDISRINSEKDIER-----L 1570
Qy 826 KTDPLPEYITGLNVYFDKSGKLIYALDMSDTIGEGQDAVGNPILNVDENEGYHTLAV 885
Db 1571 KSKLRIFA-----KSGSEI-----ETVQOELNNAQEKIRINAE--NTVLK 1609
Qy 886 ATLADYE-----GLYIKDILNSLDKIKAIROIPLAKYHRLGIFQAIRNAAE 933
Db 1610 SKLEDIERELDKQAEIKSNQOEKELLTSRLKELE-----QF 1646
Qy 934 ADRLLEPTPKGYLINEVNYRKKQOMEKMLKVPDYKTPIFNKALPNEKVDGDRAGHNINA 993
Db 1647 LDSTQKAKQSEBERRAEVAKFQVEKS--QLEKAMLETFKY--NDLVNKEQAKW----- 1697
Qy 994 ETNNSVAVTPIRSEQOLKQSQSDV--NLPTSSKNPFIYELIGVYSLCLFLVTAGKKGR 1052
Db 1698 RDEDYVAKTTSOROEIEKLAKELDNLKANSKLEKANEEDRSEIDMLMLVTDLDEKNAK 1757
Qy 1053 AR 1054
Db 1758 YR 1759

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 11:11:11 ; Search time 27 seconds

(without alignments)
3758.595 Million cell updates/sec

Title: US-10-091-007a-24

Perfect score: 5352
Sequence: 1 MKKKHKLALALTTVSVT.....VSLCLFLVTAKKGRARX 1055Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 28336 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summariesDatabase :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	331.5	6.2	1081	2 T31094	surface antigen B8
2	234	4.4	1127	2 T28317	ORF MSV156 hypothe
3	224	4.2	2748	2 S57976	nuclear migration
4	223.5	4.2	1252	2 B42771	reticulocyte-bind
5	216	4.0	2139	2 T18296	myosin heavy chain
6	215.5	4.0	1727	2 T52073	myosin-like coiled
7	212.5	4.0	621	2 A95250	cholesterol binding pr
8	212.5	4.0	2401	2 T28676	rhodopy protein -
9	212.5	4.0	2819	2 A90551	conserved hypothe
10	209	3.9	1208	2 T39068	coiled coil protei
11	209	3.9	1875	2 S38173	myosin-like protei
12	209	3.9	6713	2 B89921	hypothetical prote
13	208.5	3.9	1237	2 C97033	uncharacterized pr
14	207.5	3.9	1119	2 B70126	surface-located me
15	207	3.9	2269	2 T28677	rhodopy protein -
16	203	3.8	690	2 F98114	cholesterol-binding pr
17	202.5	3.8	2829	2 A42771	reticulocyte-bind
18	202	3.8	1940	2 A59287	myosin heavy chain
19	201.5	3.8	1365	2 T30822	lmp1 protein - Myc
20	200	3.7	872	2 S55940	telomerase compone
21	199.5	3.7	1072	2 A86827	hypothetical prote
22	199.5	3.7	3225	2 I52300	giantin - human
23	199.5	3.7	3259	1 A56539	giantin - human
24	199	3.7	1272	2 C90593	hypothetical prote
25	199	3.7	1385	2 D89824	hypothetical prote
26	199	3.7	1957	2 T38077	hypothetical coile
27	198.5	3.7	1173	2 T43527	sp8 protein - firs
28	197.5	3.7	2166	2 G70163	hypothetical prote
29	197	3.7	1790	2 S67593	transport protein

30	195.5	3.7	1300	2 I53799	CG1 protein - huma
31	195.5	3.7	1992	1 S02771	myosin heavy chain
32	195.5	3.7	2712	2 T05113	hypothetical prote
33	195	3.6	1356	2 S32763	kinectin 1 - huma
34	194.5	3.6	1864	2 F86378	protein F21J9.12 l
35	192.5	3.6	1358	2 A29360	SIR4 protein - yea
36	192.5	3.6	1839	1 OVBXK	adenylate cyclase
37	192.5	3.6	1959	2 AG1085	hypothetical prote
38	192	3.6	1026	2 C97783	cell surface antig
39	190	3.6	1132	2 H82887	hypothetical prote
40	190	3.6	1558	2 B71603	RESA-H3 antigen pr
41	190	3.6	4919	2 T31105	hypothetical prote
42	188.5	3.5	1837	2 T41023	probable nuclear p
43	188.5	3.5	4688	2 F82885	hypothetical prote
44	188	3.5	719	2 A81358	hypothetical prote
45	188	3.5	1302	1 JG6009	surface-located me

ALIGNMENTS

RESULT 1

T31094

surface antigen B8pA - Bacteroides forsythus

C:Species: Bacteroides forsythus

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #ext_change 02-Sep-2000

C:Accession: T31094

R:Sharma, A.; Sojar, H.T.; Glurich, I.; Honma, K.; Kuramitsu, H.K.; Genco, R.J.

Infect. Immun. 66, 5703-5710, 1998

A:Title: Cloning, expression, and sequencing of a cell surface antigen containing a le

A:Reference number: Z20977; MUID:99043895; PMID:9826345

A:Accession: T31094

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1081 <SHA>

A:Cross-references: EMBL:AF054892; NID:G3005672; PID:G3005673; PIDN:AA082625.1

A:Genetics:

A:Gene: bepa

Query Match	Score	331.5	DB 2	Length	1081
Best Local Similarity	23.2%				
Matches	173				
Conservative	99				
Mismatches	239				
Indels	235				
Gaps	32				
QY	63	STVDETSPDLFSDGNSN--NSSSKTESVSDPKQVXKAPVYTGASNSNDASKVPEP	119		
DB	18	TLTGATQNSGTTGPIUMSYDSGKTLATG---TGAMPDF-----NNAS--ETP	62		
QY	120	KODTAKKKTLETSTWEADFYTRGDTLVGFSKSGINKLSQTSHTVLP8-----HAAD	172		
DB	63	WHSLSQSKIQT-----VTIGGVTSVGNNAFSDCALNS-VTLPSLTAIGDHAK	110		
QY	173	G-----TQLVNASFAF-----TPDKTIAEYTSRLENGKPSRLDIDQ	212		
DB	111	GGSGLTSTIPNSVTTIGEMAFKCGSGLSITLPSNLTIGQ-SALSGCTGTLSTIPNS	169		
QY	213	KEIIDGELFNAYQLTKLTPNGYKSIQGDADVNNKNIAMVLPSELTITSDYAFAMS-	271		
DB	170	VTTIGEMAFKCGSGLSITLPSNLTIGSAFYGCALNSITLPPALTTIGSAAFKCGG	229		
QY	272	LKQVXLPDNLKVYIGELAFPDNIQGLVYLRHLIKLAERAPKS-NRIQTVPELGSKLKYI	330		
DB	230	LKSIPTFNSLTIGSAAFDCGALSTLPPALTTIGSAAFKCGSGLSITLPSN-LTTI	288		
QY	331	GEASFPD-----NNLRVMLPDGIEKTBESAFNGPDEHY	366		
DB	289	GSAAFNGSLSTIPNSVTTIGSAAFYGCGLSKITLPPDGLTTEERAF-----Y	340		
QY	367	NNQVAVRTGONPHOLAT-ENTYVNPD--KSLMRATPD-MDYTKLEDEFTYQKSVT	421		
DB	341	NGCVLTSLITL-----PNSVALIGSAAFYGCGLKSI--TLPDGTLTIEM-----G	383		
QY	422	GSNKGLOVRNRKKNLEIKQHNGLTTEIGNAPRN-----VDFQSKTLR	467		

Db 384 AFVNCAL-----TSITIPN-----SVSTIGESAFVGCALMKOVYAMDTPIIDIQRPVFR 433
 QY 468 KYDLEIKL-----PSTI 480
 Db 434 ELTSLGIRLHPAGKTYTBEAKDVWEKEFNIVEDDDFGGLQWNYDAATKTLITINPTDTP 493
 QY 481 KRTAPAFQSNLKSFEASEDLBEIKGAFPMNNRIGTLDKDKLIKIGDAFHH-INHTYA 539
 Db 494 KPMNFATPNDQI-----WGAF-QKEIKITIGDGTSGVDFAFSGCDALKS 539
 QY 540 IYLPESVQEIGRSPFRONGALHLMFIGNKVTIGEMAFLSNKLSEVNLSEQKQKTEEVQ 599
 Db 540 IYLPKSTYITIGQSAFSGCDMLRSITLPGVNTIGKAFY-DCLFELTITIPKSTYALIGQE 598
 QY 600 AF-----SDNALSEVLPNNIQTIREAFKRNHLEKVGSS 635
 Db 599 TFHYCVSLTSLTLPDALTAIGKAFYSCNMLTSTVFPKSTITIGENAF-----DGCT 650
 QY 636 TLSQITNAPQNDGDKRFGKKVYR 661
 Db 651 ALKDLTV-AMQNASIPDIGSKDVFK 675

RESULT 2

ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
 C/Species: Melanoplus sanguinipes entomopoxvirus
 C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C/Accession: T28317
 R/Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutlish, G.F.; Rock, D.L.
 J. Virol. 73, 533-552, 1999
 A/Title: The genome of Melanoplus sanguinipes entomopoxvirus.
 A/Reference number: Z20484; MUID:99102612; PMID:9847359
 A/Accession: T28317
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1127 <AFO>
 A/Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97677.1; PID:g4049717
 C/Genetic: A/Note: MSV156

Query Match 4.4%; Score 234; DB 2; Length 1127;
 Best Local Similarity 18.2%; Pred. No. 0.0032;
 Matches 211; Conservative 189; Mismatches 346; Indels 416; Gaps 56;
 QY 19 VTYSGE--VYGLERESYKQEQOTQASDDWFEEDNEKTYVS---KENSIV--DEIVS 70
 Db 108 IYKKKLYIYDLDYEEKDKKELVINIEQKNADVKNINDIKNNVNNIHSNEIITIGKETLI 167
 QY 71 DLFSQGNSSNNSKTESVSDPKOVKAKPEVTOEASNSNDASKVEVQKODTASKEETL 130
 Db 168 DIL-----NKLKLYSSDEKQILE--QIYKINNKKEIEFKKIDNVQKINKQDEL 216
 QY 131 ETSIWEA-KDFTVRGDTLVGFSKSGINKLSQTSHTLVLPSSHADGTQLTQVASFAPTPKK 189
 Db 217 NKLDESKKEFIKQDEI-----NK-----TIDK 241
 QY 190 TAIATYSTRLGNGKPSRLDIOKE-IIDE-----GEIFNAYQLTKLITIPGY 236
 Db 242 Q--EELIKKL--NDKEINFINDEKQKLDQINSKINTNENTKGVWNIYTEKKNKISNQ 297
 QY 237 KSI-GODAFVD-----NKNIAEV-----NLPESTE-TISDYAFAM 270
 Db 298 NELINKDSTIKSLDEKQKLDLSDKNNITSLYKSNKNTKINQGLLESILDPNNANI 357
 QY 271 SLKQVGLPDLNKVIGLAFEDNQIGKLYLPHHLIKLARAFKSNRIQTVEFLSGKLYI 330
 Db 358 NINEK-----SKIKLFDNDIQ-----KLANDITEQNN-KITDFPNNSTRIE 398
 QY 331 GEA-----SFQNNLRANWMLPDGLEKISEA-FTGNPDGHHNNQVAVLRTGQNP 380
 Db 399 KEKLDTEYKIDIKNNLQK--LEESYKIKIDQETEYKKNKINEYNDITIEK--NNML 453

QY 381 HOLATENTYVNPDKSLMRATPDMDYTKMLEDEFTYQKNSVTGFSNKLQKV-RRNKL-- 437
 Db 454 QKLEBKQID-----EQTEYKKNKINEYNDIIEKNNNLQKLEBKNNIND 501
 QY 438 EIPKQNGITITEIGNAFRNVDFQSKT-----LRKYDLEIKLPSTIRKIGA 485
 Db 502 KLFKANDIESNTLFPKNKINISDFKDSKSEIAKLTANTEVEBQRLKLEINIKTNELML 559
 QY 486 FAFQSNLKSFEASEDLBEIKGAFPMNNRIGTLDKDKLIKIGDAFHH-HIY 538
 Db 560 ---SDKKLSLEBQLYDSK-----NILDGIDKLYNSLKEKNDKIDEPFSNIEKDIY 608
 QY 539 AIYLPESVQEIGRSPFRONGALHLMFIGNKVTIGEMAFLSNKLSEVNLSEQKQKTEEV 598
 Db 609 NVI-----ENKIGNLDIINKI--INNQFKEYINSKI 640
 QY 599 QAESDNALSEVLPNNIQTIREAFKRNHLEKVGSSSTLSQITFNAPQNDGDKRFGSKV 658
 Db 641 DS-----KSNEL--STMEDDIFNKNQI-----ASI 664
 QY 659 VVTHNNSHMLADGERITIDPKLSTMDV-----LEKVLKITEGLD-----YSTLRQ- 706
 Db 665 TNNIENTSNKIKDNEFTISNEDSKELDEIRKYKQDFKIKDAMNTEVKSFEENTIQKD 724
 QY 707 --TTQTFREMTTAGKALSKSNLROGEKQKFLQEAQFPLGRVLDKALAKKALVTKK 764
 Db 725 IDGISKNINELTNAYDIINRKAN-----DLQ-----DK 752
 QY 765 ATKNGHILERSINKAVLAYNNSA-----IKKANVRLKEKELDLITLVBEGKPLAQATW 819
 Db 753 LNNVGESEFK-----LYNNASDLDTIOKNDKEXQ----- 784
 QY 820 QGVYLLKPLPIPEYIIGLVWYEDKSGKLIALYALMDPTIGSGODAYGNPLNVDEDNQ 879
 Db 785 -----LMEYLEKKNK--OSIEHINDIVNNFKELIK--FNNTFNKS 821
 QY 880 YHTLAVATLADYEGLYIKDILNLSIDKI-KAIRQIPLAKYTRGIFQAIRNMAAEADRL 938
 Db 822 LNELLT-----NDIDINDKIFLYKEINKI-----STNNIL 851
 QY 939 PKTYKGLNVEYPRKK-----QMEKLIKVPDYK-----TPIFN 972
 Db 852 ---KIYKNEIDNNEKLSIYENLQFINSELSIEFQGSITSHINPLNTLAGINDVLN 907
 QY 973 KALPNEKVDGPRAKGH-NINAEFTNNSVAVTPISSEQLHKSQDV-NLPQTSKNNPIY 1030
 Db 908 KL--NLKIMADTYTRGDTNIRDEIKNQISSNINISQKFNENKEDDKLISFNDKKNYN 965
 QY 1031 ELIGYSL-----CLLFVTA 1046
 Db 966 ISAGYTEYNNIEHCKLAKLYTA 987

RESULT 3

S57976
 nuclear migration protein NNM1 - yeast (Saccharomyces cerevisiae)
 A/Alternate names: protein YD8358.06; protein YDR150W
 C/Species: Saccharomyces cerevisiae
 C/Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000
 C/Accession: S57976; S19052; S17018
 R/Murphy, L.; Richards, C.; Harris, D.
 submitted to the EMBL Data Library, July 1995
 A/Reference number: S57976
 A/Accession: S57976
 A/Molecule type: DNA
 A/Residues: 1-2748 <MUR>
 A/Cross-references: EMBL:Z50046; NID:g899393; PID:g899399; MIPS:YDR150W
 A/Experimental source: Strain AB972
 R/Kormanec, J.; Schaaf-Gerstenschaeger, I.; Zimmermann, F.K.; Perecko, D.; Kuentzel,
 M. Gen. Genet. 230, 277-287, 1991
 A/Title: Nuclear migration in Saccharomyces cerevisiae is controlled by the highly reg
 A/Reference number: S19052; MUID:92079907; PMID:1745235
 A/Accession: S19052

Db 473 VKDDQSVNVYVNIQITTEBRLIVTEKORLINGIDSTINIEGALKESKG-NVEI-----GF 526
 Qy 447 --TITTEIGDAAFVAVDQSKTLEKYDELTPESTIKIGAFAPQSN--LKSEBASEDL 502
 Db 527 LEKLEIEIGKRNKLVDTYTKSINS-----TVGNFSSLFNNFDINQDFNKNI 573
 Qy 503 E--EIKGAFANNRIGTLDLKDILKIGDAAFHINHTIYALVLPESVOEIGRSAPRQNGAL 560
 Db 574 NDYENKNGEAYNEPEGSLN-----KISENARNA-----601
 Qy 561 HLMFIGNKVKTIGEMAFSLNKLSESVNLSEKOLKTIEVOAFSDNALSEVVLPPNLTQITRE 620
 Db 602 -----SENTSDYNSAKTILRLAQEKY-----NLNKEE 630
 Qy 621 EAFRNHLKEVYGSSTLSQITPNAFPDQDGRFGK--VYVETHNNSHLADGERFI 677
 Db 631 EANK--YLRDVKAYESE--RFLNNKESLDKINEMIKKEQLTVNEGHGVKQVLVENIKELV 687
 Qy 678 DDPGLSSIMVDELKVLKIEGLDYSTLRQTTOTOFREMTAGK-----ALLS 724
 Db 688 DENNLSDILKQATGKNEIQTHTSTLKNKAKTILGHVDTSAKVIGKITPELALTELLG 747
 Qy 725 KSNLRQGEKQFLQEAQFL-----GRVLDLKAIAKAEKAL-----VTKKAT 766
 Db 748 DAKLKTQAELEFESKNVVLLETENMSKNITNELDVHKNIQDAYKVALIELAHSDIDTQK 807
 Qy 767 KNGHLRESIN--KAVL--AYNNS-----AKKANVYRLKEKELDLTDIVGK 810
 Db 808 DSSKLEIENGNOIYIKVVLINOYKNISSIKSEKAVSVKIGNSKGSLSKIT-----861
 Qy 811 GLPLQATVQGVYLLKTEPLPEYVYIGLVNVPFDSKGLIYALDMSDTIGEGOKDAYGNPI 870
 Db 862 -----CSDKSVDNIIALEKQTEL--QMLRNSFTQEK 890
 Qy 871 LNVEDN-----EGYHTLAVATLADYEGLYTK-----897
 Db 891 TTNDSKLEKIKTDFESIKNALKTLEGRVNALKASSNHEVQSKSPVPALSEIKE 950
 Qy 898 ----DILNSSLDK--IKAIRQIPLAKYHRLGIFQAIRNA-----AEADRLLPPTPKGY 945
 Db 951 ETDISLNTALDELKLGKGTCEVSRY-----KLTKDYTVYKISDPTIELINTIEKNVAY 1004
 Qy 946 INEVNPKRQMEKMLKPYDYKTPILFN--KALPN--EKVDGRBAKGN--INAETNNSYAV 1001
 Db 1005 L-----AYIKKQVEDYVQDVLTLNHEFNTRQVSNHEPTNPKNSKEELTKAVDTSKTII 1060
 Qy 1002 TPRIEBOQLHKSQSDVNLFPQSSK-----NFIYELLGVSL 1038
 Db 1061 SKLKGVILIEVNTENTMTIESAKIEALYNELKNKKTSLNIEIYQTSNEVKL 1112

RESULT 5

T18296
 myosin heavy chain - Entamoeba histolytica
 C:Species: Entamoeba histolytica
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
 C:Accession: T18296
 R:Guillen, N.
 submitted to the EMBL Data Library, February 1997
 A:Reference number: Z18865
 A:Accession: T18296
 A:status: preliminary; translated from GB/EMBL/DBJ
 A:molecule type: DNA
 A:Residues: 1-213 <GUI>
 A:Cross-references: EMBL:L03514; NID:g1850912; PID:g1850913; PIDN:AA84805.1
 C:Genetics:
 A:Gene: mhca
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 F:91-780/Domain: myosin motor domain homology <MMO>

Query Match 4.0%; Score 216; DB 2; Length 2139;
 Best Local Similarity 20.2%; Pred. No. 0.056;
 Matches 260; Conservative 168; Mismatches 494; Indels 342; Gaps 58;

Qy 3 KKHLLTALATLTVSVVTVYVTSQEVYGHRESVYKQETOSASEDDWPEEDNERKTIVSKEN 62
 Db 883 KRBAENALALSAFAKTEGELAEKIQDLEDKISLESLTSAEELDK--QELNLTKIENLEDK 940
 Qy 63 STVEFVSDLPFSDGNSN-----NSSKT-----ESVSDP 92
 Db 941 EELKETIDNKGDLKDKSLKGEDLEVEITELNSQITLNAVTDKKTIAEMOESIDEKE 1000
 Qy 93 KQYPAKAPV---TQASNSNSDASKVEVKOTASKEKTELSTWBAKDFVTRGDTLV- 148
 Db 1001 DEITTKGDIKLEEDKDLEQDRADVASATKDIKAKKITTICEPAKBIATLBELE 1060
 Qy 149 ---GFSKSGINKLSQTSHTLVPASHADGTLTOVASPAFTPDKTAIA-----YTSR 198
 Db 1061 DEENKNDLNLQQT--LKGTEPKSLAAQVAANKKASDERDTSQMLENEKLTQKN 1117
 Qy 199 LGE---NGKPSRLDIOKEIID--EGEIPNAY-----QITKLITPNGY---236
 Db 1118 LTKTKADLEKKISGLKQDYEDLEDKNNKIEGDLRNAQRKTELDEITTKGADVSQYLOKQ 1177
 Qy 237 -----KSIQDAFVFNKNIAE-----VNLPEST--ETISDYAFAMSLKOV 275
 Db 1178 KEYESQIAKQEEKEKAIQNDVKNKEKTIYKEKELFISLOEKDETEVEKDAKKKEI 1237
 Qy 276 -----KLPPNL-----YVIGELAFVNOIGKLYLPHLI 305
 Db 1238 EKEMKALQEEKENESSKSTEDKCKLELDLKTQKDLDMT--ADNE-----1284
 Qy 306 KLAERAKSRNIRQVVEFLGSLKTYIGEASF-----QDNLRVNMVLP--DGLKTESAF 357
 Db 1285 KIKAKA--KDLQAQNLNEQVQDNEKAVADAEILINKKKAOSDEINLKALEALTYAKSVYE 1343
 Qy 358 TGNPGEHYNNQVLRFRFGQNPQHLATENTYVNPDSKLMPATPMD-----404
 Db 1344 SKAQDSE--NEKALSEIIOANEK--NIQADIRKATADLQANEKKAVEAQR 1395
 Qy 405 -----YTKWLE-----EDFTY-----QNSVTGSNKL--QVRNKN 436
 Db 1396 DKLVAQNKQKTKLEIKARDENTYKVENYEKYLKREKADLEBANENLDLEKQDRNKE 1455
 Qy 437 LEIKQNGITITTEIGNAF--RVNDPQSTLAKYDEEL-----KLPSIR 481
 Db 1456 KOYKVLBELEKETQDKNAALIAEKDSIFTAQKSDADLELNTKEVHEDEVAALQNTIT 1515
 Qy 482 KI-----GAFAPQSNLKSFEASEDLLEIKE-----GAFMNNRIGTLDLKDILKIGDAF 532
 Db 1516 KLTNDNGAEBELNEIR--KADKQKKISLEBQVNELESRPVGTGNADNEIETLRDA-- 1572
 Qy 533 HINHTIYALVPEVQEIGRSAPRONGALHLMFIGNKVY--TIGENAFPSN-----580
 Db 1573 QIADLNKALIEKGVQNNQLQATNKEKAKDNDLTSKIEITENEMKKLENAKKRLQDQDE 1632
 Qy 581 -----KLESVNLSEKQKLTIEVQ--PSDNALSEV-----LPPNLTIREBAF 623
 Db 1633 ADKAVSEQTIRKGGIEEVKULTTIRIQLKQINAPSSVAAQBEKQRLSEDLAEKEQL 1692
 Qy 624 KEN-----HLKEVGSSTLSQITENAPDQDGRFGKQYVYVTHNSNM-----668
 Db 1693 QERTTAANAABERKKIQAEIDQVKNLMDVYNQRE--KLVAKSENDAEISLKEEKA 1749
 Qy 669 IADG--ERFLIDPDKLSTWVDEKVLKTIIEGLDYSTLRQTTOTOFREMTAGKALL--SK 725
 Db 1750 LEDEIEKITTDDNNKLSSEIISIDRKYKNAL--LDSKSDSVSMKEKQDELKVTQDLATEK 1807
 Qy 726 SN---LR--QGEKQFLQEAQFLGRVLDLKAIAKALYKKAATKXGHLIERSINRAV 780
 Db 1808 KNAHTTMRKIGLEKEAAEVQVRLALQKNDLDAQEKA-----KATXQYRABDGLKSLM 1863
 Qy 781 IAYNNSAIKKNVY--RLEKELDLTLTVGEGKPLAQ--ATWQGVYLLKTPLEPEYIIG 837
 Db 1864 NELD-----DVKQQLDKAQDLDLADKEDLATTLDQKKTLYVQKSVFDSIRIQEMQEOQLD 1916

QY 1039 CLEFLV 1044
 Db 1059 LKTYIV 1064

RESULT 9

conserved hypothetical protein MYPV_3130 [imported] - Mycoplasma pulmonis (strain UAB CT A90551)
 C/Species: Mycoplasma pulmonis
 C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 R/Accession: A90551
 R/Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.; Nucleic Acids Res. 29, 2145-2153, 2001
 A/Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis
 A/Reference number: A9512; MUID:21267165; PMID:11353084
 A/Accession: A90551
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-2819 <KUR>
 A/Cross-references: GB:AL445566; PID:G14089727; PIDN:CAC13486.1; GSPDB:GN00153
 A/Experimental source: strain UAB CTIP
 A/Genetics:
 A/Map position: MYPV_3130
 A/Genetic code: SGC3

Query Match 4.0%; Score 212.5; DB 2; Length 2819;
 Best Local Similarity 18.2%; Pred. No. 0.12;
 Matches 246; Conservative 194; Mismatches 444; Indels 467; Gaps 56;

QY 1 MTKGHLKTTALA---LTTVSVVTVYSGEYVGLERBSV---KORQTOGASDDWFEED 51
 Db 3 MKKKIILIGLASMAGIITTSAILPK--IGKGDVAVVKNKKDSETSGSVSDPDKN 60
 QY 52 NEKTKVSKENSTVDETVSDLSFGDGNSSSKTESVSPKQVPPAKPEVTQEAAGSN 111
 Db 61 TKSSPDKKANQ---NKKKSQSRSSESD---LKNKNSN 96
 QY 112 DAKVEVPKQDTASKKETLETSTWEAKD---FVTRGDTL-VGF-----SKS 153
 Db 97 NFPSGPIIDNNSTISNDKRL---IKDKIRFVTFGDSIAGNNAKLDVPEEYDPNTK 150
 QY 154 GINKLQSTHLVLPASHADGTOLQVASFAPTPDKTALAEYTSRLGE---NGKPSR 207
 Db 151 QITGLASYSTIADYINDIDNPKLESFNFNFAFS---GTLKDMWDYFNKQOTVFPN9TKSE 207
 QY 208 LDDIDQKRIIDEGEIRFNAVQUTKLTITNGYSGIGQDAVDNKNIAEVNLPESLETISDYAF 267
 Db 208 -DLSQREDF---IKRLT-----NSNLLILNLGAN-----DF 234
 QY 268 AHSLSKQVRLPDLNKKVIGELAFPDNOIGKLYLP---RHLL-----KLABRAFKSNRI 317
 Db 235 MSIAFEKFEKDVLSLMKKDASTNDILG-VFLPIISITHENGIRYKELISIRKYNK 293
 QY 318 QIVEFAGSKL-----KVIKESAFQDN---NLARNVLPDGLKTESAFTG 359
 Db 294 VTINLIGYPMPALKEFMVNVKLIGESYKIGNDTLGLFNLFINSSIKDQUDVLEANNFVW 353
 QY 360 NPED-----EHYN-----NQVVLKRTGQNHQALTE-NTYVNPXK-SLWRAT- 400
 Db 354 AYDDPMLKHKNDFTDVAFDIHENEIGYKMAQEIFLKLALIDSKNNPSEINSSNNANY 413
 QY 401 PDMDYTKWLE-EDFTYOKNS-----VTGFSNKGIGQYRKNKNIIEPKOHGINTITEIGDN 454
 Db 414 LSYDHGKFSRLFEFKKAKDSEIITTYLVGVNNSALFSGNDKRIALYESKFTDNYSEI--- 470
 QY 455 AFENVDPQSKTLKLYDEIEIKLPSITRKIGAPAFQSNLKSFEASDEIEIKGAPMNR 514
 Db 471 LFRYSYDI-FKTLKSAYE-----FTINSIFKTLDPEDL---KKFLPANN 513
 QY 515 IGLDLIDKDKLIKIGDAFHNHIVAYLVPSVOEIGSARQNGALHMTGKVKVITGE 574
 Db 514 ANYQKLFELTKRFPVHNINRSIRDTLTLLDNNQAGIQETELSHLMTTFPKK-----E 568

QY 575 MAFLSNKL-----ESVNI-S-----RQOKLTIEVQAFSDNALS 607
 Db 569 ISNLDNYLGLRLEFFSSSPYVDETNLSKPDIIKIKLKNFLQKSEKTLTLLSKETAFLIO 628
 QY 608 EYVLPENIQTIREEAFKRNHLKRVKSGSTLSQITFNAFPQNDGKPKKVVVR----- 661
 Db 629 NKILSNLSEL-----VLRIVSSEKTFALI-----EKILDDIFDNKAVYQYKXIE 674
 QY 662 -----THNNSHMLADGERFIID---PD----- 680
 Db 675 DFLFAFVNNEKAITDFVRFVNLKAVKENFALVSGISQPLNSNEVISTQTSQAFY 724
 QY 681 KLSSTWVD-----LEKYLK-----IIE-----G 698
 Db 735 EVSKILVNIQVTEBELLEKLVKFEIKSSAQKLSNPKFENILEKLOKSIIITPGFTND 794
 QY 699 LDYSTLRQTTQT-----QPREMTYAGKALLSKSNLRQGEKQFLQEAQFPLGRVLDK 751
 Db 795 LGMSPFLQYVKSLSLSVQROKFTSSLSKEVVSFVFKSENVKKNKLSSAGTYLSNTELFK 854
 QY 752 AIAKAEKALVTK--KATKNGHLERSIN-----KAVLAIN-----NSAIKKA 791
 Db 855 ENSDVEKGIFFSHILNFEKTPPELLNKIIDVIVNSWESIDQYNDAFBEVLNFKLTNSWFKT 914
 QY 792 NVKRLKEKEDLLTDVEGKGPLAQATWQGVYLLKTPLPBEYVIGLNVYFDKSGKLI-- 849
 Db 915 YVK-----DIPSSVNSSEIQA-----NILSTLANKLASQ 944
 QY 850 VALMSDTTIGEGOKAYNPPIINVEDNEGHYTLAVATLADYEGLYKIDILNSLDK--- 906
 Db 945 AGFELTNSVESIKNIFTNLIKLIADSN-----TIDALVD---TFVNSKTLLEIDKKNF 995
 QY 907 ---IKAIROIPLAKHRLGIFQAIRMAAEABR----- 936
 Db 996 IASLNKIOIFALSDNFFSVFALISSNKLKQKAFVQYKODITVNLKELLEKEKFPNLI 1055
 QY 937 ---LIPKTPKGY-----LNEVNYKKQMEKRLK----- 963
 Db 1056 YSFMFMKTEISELPENNPEALNKTTITIRNALKKNLDDFINKMMDITFDMLDKQKLENYG 1115
 QY 964 -----VDYKTPIFNKALPENK-----VDGDRAKGNINVAETNSVA 1000
 Db 1116 QVILGLIKONSDALARKVQAFIDQVNVBANFQVLKVLVSGQKFKMGELINADERTNKIS 1175
 QY 1001 VPIREBQOLKHSQSDVNLQOTSCKNFTYE 1031
 Db 1176 TDLTKILQVFNESNFAKSLALATKDSLVID 1206

RESULT 10

coiled coil protein - fission yeast (Schizosaccharomyces pombe)
 C/Species: Schizosaccharomyces pombe
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 R/Accession: T39068
 R/Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Connor, R.E.; Wood, V. submitted to the EMBL Data Library, August 1997
 A/Reference number: Z21825
 A/Accession: T39068
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1208 <MDR>
 A/Cross-references: EMBL:281317; PIDN:CA03608.1; GSPDB:GN00066; SPDB:SPAC669.06C
 A/Experimental source: strain 97zh-; cosmid c669
 A/Genetics:
 A/Map position: 1

Query Match 3.9%; Score 209; DB 2; Length 1208;
 Best Local Similarity 19.4%; Pred. No. 0.051;
 Matches 220; Conservative 170; Mismatches 367; Indels 378; Gaps 52;

```

QY 20 TYSGVYGLERESVKSQOQSA-----SEDDVEEDNEKRTVSKNSVTDEVSDFLS 74
Db 93 SFDNYSYGL---SAISKQATQEBALISIQGDSY---DVSKITDLSK-NSEIDHTDGL-- 143
QY 75 DGNSSNSSKTESVYSD-----PKQVPKA--KPEVTOASN 108
Db 144 PANALITIRBQEKYLEKVSRENFGRIKIVCLEKRLSMAPEQIKAVKONVELHABRAN 203
QY 109 SSNDASKVEVPKQDTASKKETTLETSTWPAKDFVTRGDTLVGFSKSNKLSQ-----T 161
Db 204 LQLOKRTRESLLOKSEDKNFLE---EKVYLSK-----VDVQSQNVKVFY 248
QY 162 SHVLPBSHAGDTQUTQVA--SFAFTPDKTAIAEY-----TSRLGKNPSRLDIDOK 213
Db 249 ERIFLEALTEKVRKESKISTEMEEDSKNKEVDEYETRIQLOKRLDELSE--ELDVAQ- 305
QY 214 EIIDEGEIFNAYQUTLKITPNGYKSIQDAAFVNKNIAEVMPE-----257
Db 306 DLUTEKE---DEIATLKQJEBKENSASSAFENENSUYHLOEDYAILQAKCDEPADRI 361
QY 258 ---SLETTSDYAFAMSLKQVYKLPDNLKYIGSLAFPDNQGKLYIPRLIKLAERAF 312
Db 362 QVLTADLEKEREKNOIMESEASIGILDSMOV-----H--TLOEQH 400
QY 313 KSNRIQVEFLGSKLKVIGE--ASPDNNLRNVM,PDGLKTESAFPTGNGDEHYNNQV 370
Db 401 KAN--EEIEFLHDQISRNESGKNFEDIMLOPRSLSEEDVLESLOQ---LEDNNSL 454
QY 371 VLKTRT--QONPHOLATENTYVNDKSLMRATPDMDYTKMBEDFTYQNRSVTGESNKLQ 429
Db 455 RUMTSILGNQJESLRTQREIDEENHMLRLAKNSDKALAE-----TIRLQ 502
QY 430 KYRRENKUIEIPKONGITITTEIGDAPRN-----VDFQSKTLARKYLE 472
Db 503 EV--TKELLETIRMKNSNDLNEIHDLRNEEGITLKIDSITKEKDLINELBOKRKSXYEVN 560
QY 473 EIKLESTIRK-----IGAFAPOSNNLKSEFASSEDLEIKEGAFMNNRIGT 517
Db 561 VSELNGTIDEYRNLKQKEETYNEMMAFYQKNDLRFHESINLQJOREKELISN---- 616
QY 518 LDLKDKLIKIGDAFHINHIAIYVLPESVOEIGRSAPFONGALHMFIGNKVKITIGEMAF 577
Db 617 LEKRLVLI-----SSIRFVAMLEKERESEIKKYLISGN-----648
QY 578 LSNKLESVYLSF-----QOKKIT--EYQAFSDALSIVLP,PNLQITREARAFRN 626
Db 649 -AKDLNNTNMEIINDKISVLQROLTVYDELDVSEEREERAVAGQKLSASFELMSNEK 707
QY 627 HLKEVKGSSLSQITFNAFDDNDGDR-----FGKTVV---VRTNNSHMLADGERFI 676
Db 708 QALTEKYSKLMEL--INA--QMLDRREBELSELSKULFEERKISGSDNDIEKKKEINV 764
QY 677 ID--PDKLS-----STWVLEKY-----LKIIISGLDYSTR- 705
Db 765 INSEIADKLQAIRHLESQMEDLDKLVHNLNGIEBEANIENNAVKKRLCILMGCDSVSI 824
QY 706 -----QTOOTOFREM-----TAAGALLSK 725
Db 825 LQIVQOIEHFVNOQIOTIRSLKQELRHDFVQSGKKQOELSSFEKGCITETKCDILAQ 884
QY 726 SMLROGEKQKLOE--AQFPLGRVD-----LDKAIKAEKALVTKKATK 767
Db 885 RRRANVSEKNDLENNAQKFFSPDRKNGYLPSEHTSKIEYLEKTIEDLKAL--QDELK 942
QY 768 NCHLLERSINKAVLANNNAIR--KANVKELEKELDLTD-----LYEGKG 811
Db 943 NNNLMDIDIS---SYNKQITTKQEKIKVLEBERSILIDELBSYESNOFNYYONNVODKN 998
QY 812 PLAQATMVQGVYLLKTPPLPEYVYIGLVNY---FDSGKLYIA--LDMSDTIGEGQKDAY 866
Db 999 ELIEER-----LKEIQLKELEYNNHFMQALIMTSNVTDESQMLKTLIRREL 1044
QY 867 GNPITINVDENEGYHTLAVATLADYEGLYIKDILNSSLDKIKAIQIPLAKYHRL 921

```

```

Db 1045 QSKTNNDH-----LSITLERNREKYEKSL-----LDDYNGLR-----ARYKNL 1082

RESULT 11
S38173
Myosin-like protein MLP1 - Yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKR095W; protein YKR415
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C;Accession: S38173; S40647; S31207
R;Balladon, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo
submitted to the Protein Sequence Database, March 1994
A;Reference number: S38158
A;Accession: S38173
A;Molecule type: DNA
A;Residues: 1-1875 <BAL>
A;Cross-references: EMBL:X28320; NID:9486586; PID:9486587; MIPS:YKR095W
A;Experimental source: strain S288C
R;Bou, G.; Esteban, P.F.; Balladon, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.;
Yeast 9, 1349-1354, 1993
A;Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chro
A;Reference number: S40644; MUID:94205265; PMID:8154186
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:9450550; PIDN:CAA51948.1; PID:9450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S31207
A;Molecule type: DNA
A;Residues: 1-300; 'A', 302-1875 <KOE>
A;Cross-references: EMBL:L01992; NID:9171958; PIDN:AAA34783.1; PID:9171959
C;Genetics:
A;Gene: SGD:MLP1
A;Cross-references: SGD:S0001803; MIPS:YKR095W
A;Map position: 11R

Query Match 3.9%; Score 209; DB 2; Length 1875;
Best Local Similarity 19.2%; Pred. No. 0.098;
Matches 221; Conservative 194; Mismatches 375; Indels 362; Gaps 55;

QY 32 ESYVQEOQSAF-----DWFEEDEKRTVSKNSVTDEVSDFLS-----ESDGSNN 79
Db 107 ESVENKMKLSSELFPYRKLDLITBEKKERQSQOQRTKLIDRLKEITELVRYENNRSN 166
QY 80 NSSKTESVSDPKQVPAKPEVQASNSNSNDA--SKVEVPKQDTASKKETT--ETSYWE 136
Db 167 SECKLRSTIMDL-----TKQGYITNDLSNRTELARK--TQELTLLQSNNDWL 214
QY 137 AKDFVTRGDTLVGS-----KSGINKL-----SQTSHVLPESHADGTQUTQV- 179
Db 215 EKEKRSKNEQYLSYRQKTDKYLIDIRNELNLRNDFQWERNNNVYLKQKNNELSSQEK 274
QY 180 -----ASFAPTPDK-----TAIAYTSRLGKNPSRLDIDOK 213
Db 275 LIETKGLSDLSNSEKQESASMSLKQRLVDLBSQINAVKEINSIRLNATKAYTADSK 334
QY 214 EIIDEGE--IFNAYQUTLKITPNGYK-----SIGDAAFVNKNIAEVMPE,PSLETSI 266
Db 335 KQTPENEDLKELOUTKTEKKAQCEKEGCLRLSITDEADENENL-----SAKSSSDF- 386
QY 267 FANMSLKQVYKLPDNLKYIGSLAFPDNQGKLYIPRLIKLAERAF--SNRIOT--VEF 322
Db 387 -----IFLKKQILK--ERRTKHLONOLETFLVE 413
QY 323 LGSRLKVGASPDNNLRNVM,PDGLKESSEAFPTGNGDEHYNNQ--VVLKTRTQGNP 380
Db 414 LEHKVPIIN--SFKERT-----DMLNELNNAALLLEHSHNEGNAVYKELNNAQ 461

```

```

Qy 381 HOLATENTYVNPDKSLMRATPMDYTKMLEEDPTQKNSVTGFSNKGLOK--VRNKNLE 438
Db 462 KLVBCENDLQTLTKQ-----RDLGCRQIQ--YLLINVSNSKGLRKEEIIQIENI- 512
Qy 439 IPRQHGIIITTEIGDNAP---RNVDFOSKTLRKTYDEILCPSTIRKIGAPAPQSNLKS 495
Db 513 --MOEDSITTESDQKVTTERLVEF--KNIIQLQENRABLKVVENL-ADKLESKEKKS 567
Qy 496 FEA-----SDLEIKEXGAFMNNRIGTLDKDKLIKIGDAFHNHLYALVLPESVOIG 550
Db 568 KQSLQKTESSTVNEAK-AITLTKSEKMDLESRIE-----LQKELEEL- 610
Qy 551 RSAFRONGALHMEFGKVTIGEMAFSLNKLSEVNLSEOKOLKTIEVQAFSDNALSEVV 610
Db 611 -----KTSVPNEDASYSN-VTIKOLTEKRDLESQVQ----- 641
Qy 611 LPPNLQITREAFGRNHLKAVKGSSTLSQITPNAFDQNDGKRFKKVAVVTHNNSHMLA 670
Db 642 --DLQT-----RISQITRESTENMSILNLEIDIDLYDSKSDISIKLGEKESRILA 689
Qy 671 DGBRFTIDPDKLSSTWVDEKVLKIIIGLDYSTLRQTTQIQ--PREMTTAKALLS----- 724
Db 690 E-ERFKLISNTLITLTKANDQLRKRFYLONTIILKQDSKTHEITLNEYVSC-KSKLSIVET 747
Qy 725 -----KSNLR-----QGEKQKFLQEAQFELGR 746
Db 748 ELNLTKEQKLRVHLKQKQELNKLSPKQSLRIMVTQQLQTKEREDLLEET-----R 802
Qy 747 VDLDKALAKAEKAL-VTKKAT--KNGHL--LEBSINAVLANNASAIKKNVRLKEELD 801
Db 803 KSCQCKIDLEDEDAISELKESISQODHIKQLEEDNNSNIEMVQY-----KIEMLKDYE 856
Qy 802 LELTDLVEGKGLAQTAVQVYLLKTEPLPEYVYIGANVYFDSKGLIYALMDSDTIGEG 861
Db 857 SVITSVDSK-----QTDIEKIQYKKS-----LEK----- 881
Qy 862 QKDAYGNPILNVDEDNQGYHTLAV--ATLADYEGLYIKDILNSSLDKIK-----AIRQIP 914
Db 882 -----EIEEDKIRLHTYVNMDETIND-----DSLRLKELESKINLTDAYSQIK 924
Qy 915 LAKYHRLGIPQALINMAAEADRLLPKTPKGYLNEVPYRKQKMKLKPVDYKPIFN-- 972
Db 925 EYK---DLVETTSQSLQNTNSKLDSESKDPTQNTQIKLDTBKTSLEDKISILKQEMENLN 980
Qy 973 -----KALPNEKYDGBRAKAGHINAVETNNSVAVTPIRS--EQOLHKSQSDVN--LP 1020
Db 981 NEHDLQKKMEKERADEPRK-----RISLQNNNKEAVEAVKSEYSEKLSKIQNDLDQQTIV 1035
Qy 1021 QTSKKNPIYEI 1032
Db 1036 ANTAQNNVYEGEL 1047

```

RESULT 12

```

B89921
hypothetical protein ebha [imported] - Staphylococcus aureus (strain N315)
C/Species: Staphylococcus aureus
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C/Accession: B89921
R:Kuroda, M.; Ohts, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Oi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramoto, K.;
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A/Reference number: A89758; MUID:21311952; PMID:11418146
A/Accession: B89921
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-6713 <KUR>
A/Cross-references: GB:BA000018; PID:513701232; PIDN:BA842527.1; GSPDB:GN00149
A/Experimental source: strain N315
C/Genetics:
A/Gene: ebha

```

Query Match 3.9%; Score 209; DB 2; Length 6713;
 Best Local Similarity 17.6%; Pred. No. 0.65;
 Matches 199; Conservative 207; Mismatches 406; Indels 318; Gaps 46;

```

Qy 29 ERESVKQOTOSASEDDMFEEDEKRTKNSKNSVTDETVSDLFSDGNSNNSKTE-- 86
Db 5190 DKKEAVDQALQAAQSTITDPTGNSNANKDAVEQALTKLQEKVNL--NGNERVAEAKTQAK 5247
Qy 87 -----SVWDDPKVPRKAP--EVTQASNSNSNAKVEVVKQDTASKKET 129
Db 5248 QTTDQTLTNADQIAKQKQITDQATKQPLAEIVDQATQLOSMQDQQAQVNEHANEQT 5307
Qy 130 LETSTWBA-----KQVTRGDTLVGFSKSGIK--LSQTSNHLVPSHADGTQLTVAS 181
Db 5308 IDYTOADSDKQKAYKQALDAENVL--KQNAKQOVDAQLONLN-----AK 5352
Qy 182 FAFPPDKTALAEYTSRLGNGKSPSLDIDQ-----KEIT 216
Db 5353 QALNGDERVALAK-----TNGKH--DIDQLNALNNAQDGKGRIDQSDNDLQIQIIV 5403
Qy 217 DGEIINAV--QLTKLTPNGYKSGIGDAFVDN-----KNAEVLNPSLETISDYAFAM 270
Db 5404 DEKALNPRAMDQLSQETITGEGRTKSTNYNADTVQKQYDAVDPAKQALDKSSQNL 5463
Qy 271 SLKQV-KLPDNL-----KVIGELAFPDNQGKLYLPHLTKL--AERAFKSNRIQVTEFL 323
Db 5464 TAOQVILKNDVAVNAAKKALNGEERLNRRKAEALQRLDQLHNNNAQQLALQIINNAETL 5523
Qy 324 GSKLKVYIGASFDNNLRNVMPLPDGLEKISEAFNGPGBEHYNNQVVLATRTGQNFQOL 383
Db 5524 NKASRAIRNATYKLDNM-----GAVOQYIDEBHLGVISSSTNYIN----- 5562
Qy 384 ATENTYVNPDKSLMRATPMDM-----YTKMLEEDPTQKNSVTGFSNKGLOKVR 433
Db 5563 ADDNLKANYDNAIANAHELDKYQGNALAEABOLKQITIDQANALNGQNLANKADKA 5622
Qy 434 N-----KNLEIPKQNGITTEIGDNAPFNVDPOSRTKRYDLBEIKLPSTI 480
Db 5623 NAFVNSLNGLQOQODLAHKAINNADTVSDVTIVNNQIDLANDMETIKHLVNEIRPAE 5682
Qy 481 RKIGARFQSNLKSPPASADLEIKEGAMNRRIGTLDKDLKIKIGDAFHNHLYAI 540
Db 5683 QTVYQVADNADNATNF--DARLANTLLNSDNTVNDNGALQAVNDAIHINIGDQRL 5739
Qy 541 --VLPSVOEIGSAFRONGALHLMFIGNKYVTIGEMAFSLNKLSEVNLSEOKOL--K 594
Db 5740 QDAKDKALQSLNQA-----LANLK-----EIEASNADQDQLIAKAK 5777
Qy 595 TIRV-----QAFSDNALSEVVLPPNLQITREAFKRNHLKEVKGSSSTLSQITENAF 645
Db 5778 AELLANSIINNKKATNSQAVSQV-----QTAGNHAIQVHAIEIPKAKI----- 5822
Qy 646 DQNDGDKRFGRKVVYRTHNNSHMLADGERFTIDPDKLSSTWVDEK-VKTIIEGLDSTL 704
Db 5823 --DANDVDKQV-----QALIEIDIRNPMILTDEKQALK-----DRI 5857
Qy 705 RQITQOTFRREMTTAGKALLSKSNLRQGEKO--KFLQEAQFLL-GRVLDKALAKAEKALV 761
Db 5858 NQIILQGHNDINNA-----LTKELIEQAKQAQLAQLDQIKQLVAKEDAKQDVDAQVQALI 5913
Qy 762 TK-----KATKNGHLLEBSINAVLANNASAIKKNVRLKEKELDLITDLVEGKGPLAQ 815
Db 5914 DEIDQENNLTDKEKQALKDQINQ--ILQOQHNGINNANMTKE-----EIEQAKQAQLAQ 5963
Qy 816 ATWQGVYLLKTEPLPEYVYIGANVYFDSKGL--LIYALDMSDTIGEGKQKQVYGNPILNV 874
Db 5964 A-----LKEIKDLVAKAKENAKQVDKQVQALIDEIDQNPULTKQKQALKRINQIL 6015
Qy 875 EDNDEGYHTL-----AVATLADYEGLY-IKDIINSSLDKIKALRQPLAKYHRLG 922
Db 6016 Q--QGHNDINNAKMERIEQAKQAQLA--QALQDIDKIDLVAKEDAKNMIKALANK-- 6066

```

Query 923 IFOAIRMAAEADRL.PKTPKGYLINEVYNYRKKOMEKNLKPVDYKPIENKALPNKXVDG 982
 Db 6067 -----RDQINSVPDLTPKQAKAKEID-----EA 6091
 QY 983 DRAAKGHINAEITNNVAVTPPIRSEOOHKSQSDVNLPTQSSKNFIYEI 1032
 Db 6092 EKALQVVENAQITD-----QLNRG---LNLGLDIDIRHTHWEV 6127

RESULT 13
 C97033
 uncharacterized protein, probably surface-located [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: C97033
 R:Noiling, J.; Brelton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: C97033
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1227 <KUR>
 A:Cross-references: GB:AB001437; PIDN:PAK79054.1; PID:G15023995; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 A:Genetics:
 A:Gene: CAC1080

Query Match 3.9%; Score 208.5; DB 2; Length 1227;
 Best Local Similarity 18.9%; Pred. No. 0.055;
 Matches 221; Conservative 164; Mismatches 363; Indels 419; Gaps 54;

QY 3 KKHILKTALALTTSVVTYSGEYVGLREESVQEOQASAE-DMDFEEDERKTNYSKE 61
 Db 5 KNILLVFSISVTTSAVLISKPALVAYAADNSVYSSISNSNEIIVKGEKEQVEKAVNKN 64
 QY 62 NSTVDETVSDLFSNGSNSSKTESVSDPKQVPAKPEVTOEASNS-----NDAS 114
 Db 65 N-----DSNSKVSNNENOVSNKNSNPXSSSEIIGINKNNVNLQVQONKNS 112
 QY 115 KVEVPKQPTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHTLVLPESHADGT 174
 Db 113 VLAASNVDDEVKK--INGNVQTSYDIAIGETKV-----144
 QY 175 QLTQVASFAPFPDKTAIAEYTSRLGKNGKPSRLDIDQKEIID-----EGEIPNA 224
 Db 145 -----KEDTLDIINKALVDARSAGTDLSEVEIYDI 175
 QY 225 YQLTFLTPNGYKSGDPAFYDNKNIAEVLNPESELETISDYAFAMSLKQVLPDMLKVI 284
 Db 176 VSGTKEAL-----QEA-----RNINDGVANVSQYT-----LL 203
 QY 285 GELAFEDNOIGKYLPHHLIKLARRAKSNRIQVEVLESGKTKVIGASFOQNNLRVW 344
 Db 204 G--ATFVNDAN-----LDSVVKYFHHKRYATV-----TKFKDV--ATKTSNAKNI- 245
 QY 345 LPDGLKIESEAFNGPDGDEHNNOVLRTRTGQPHOLATEENTYVNPDKSLMRATPMD 404
 Db 246 -----NN-----GQG-----GDT 254
 QY 405 YTKMLBEDFTYQKNSVTGFSNKGLOKVRNNKLEIPKONG--ITITEIGN-----454
 Db 255 YLA-LE-----VSGVTQPYLDLV--NKNIYVEKQNGKHDTITTEISDAASTIARI 302
 QY 455 -AFKNVDFQSKTLKRY-----DLEIKL-----PSITRKIGAFAPQSN 491
 Db 303 NTAIDNNMAGVATLLDYOALIGANNVPLQHLVADVNSLADQKNGDVSEALDINTMTYIN 362
 QY 492 NLKSFASHEDEIEIKGAFMNNRIGTLD-----LKDKLKIGDAAFINHIVAIYLP 543
 Db 363 NINSGVGTED-DYINSHAVDSNE-GNIDDIILANITIEKTKAKGD-----LITP 410

QY 544 E---SVQEI-----GRSAFR--QNGALHLMFGNKVKTIGEM-----AF 577
 Db 411 EVANVYKEVKTLLDFYNHAAQOTLLQYKQVDPNPAQVQODVATLSDMLKTRDCKTLKA 470
 QY 578 LSNKLESV-----NLSEQOKLTIEVQAFSDNALSEVVLPHLQITRBAFRK 625
 Db 471 LQDKIDSILNSKINSKINSIGNIDVYSKQTEVAVDASKLEAVND-----DIKIKDKGSD 525
 QY 626 NNLKEVKS--STLSQITFNAPONDGDRFGKVVVFRTHNSHMLADGERFLIDPDKLS 683
 Db 526 LTIQIRBDSVKTIDYIN-STNSVSGKSDSVSDYITIGDVEINIE--FVNEKIKS 581
 QY 684 STWVLEKVLKIEGLDYSITLROTTQTFREMTACKAL-----LSKSNR-----729
 Db 582 GITITENIKVYIE-----PIVQISEVYRIYGVGVVDYKTLGINNVNDNNIYINAE 636
 QY 730 -QGEKQKFOEAF-----FLGRVDDLKAIAKEKL-----VTKKARKNGHLRSTIN 777
 Db 637 LKKKKQVKIQDITRVNNTINNIDVINKIGAGAVLSDYFNIGIT-----DYV 684
 QY 778 KAVLAYNNSAIKKANVRLEKEIDLLTDLVEGKPLAQTMYQGVYLLKTPPLPEYITG 837
 Db 685 QDILDVYNDLKIQNV-----KQVDILIEVKAISSEYALM-----RINGEAVTDFKA 725
 QY 838 LNVYFDKSGKLIYALDMSDTIGEGQDANGPILNVDEDESGH-----TLAVATLADYE 892
 Db 736 LGTLTINDIGLLVA--TTDLONNKYKA--DEVIAVQAOLEIRALMQINLGRATPDYN 792
 QY 893 GLYIKOILNSSLDKIKAIHQIPLAKYHRLGIFQAIRMAAEADRLPKTPKGYLN--EVP 950
 Db 793 TLGILIDVNDLILYVNAIDQ-----GKNVNVNDEV- 822
 QY 951 NYRKQWEKNLKEVDYKTPIFNKALPNEKYDGRAA--KGNINAEITNNVAVTPIRS 1006
 Db 823 ---KAEIKKNIQ-----IYNALL--KIDSGSATIDYRTIGITFVIDYNISYVIRI 869
 QY 1007 EQOLHKSQSDVNLPTQSSKNFIYEI 1033
 Db 870 K-----GMNIPQVSDAKRFINTIIL 888

RESULT 14
 B70126
 surface-located membrane protein 1 (lmp1) homolog - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 11-Jan-2000
 C:Accession: B70126
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whi
 son, D.; Peterson, U.; Kerlavage, A.R.; Quakebush, J.; Salzberg, S.; Hanson, M.; Vu
 g ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:9805943; PMID:9403685
 A:Accession: B70126
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Cross-references: GB:AB001131; GB:AB000783; NID:92688098; PIDN:MAC6595.1; PID:92688
 A:Experimental source: Strain B31
 C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat
 F:742-774/Domain: tetratricopeptide repeat atypical <TT1>
 F:775-808/Domain: tetratricopeptide repeat homology <TT2>
 F:809-842/Domain: tetratricopeptide repeat homology <TT3>
 F:843-876/Domain: tetratricopeptide repeat homology <TT4>
 F:911-943/Domain: tetratricopeptide repeat homology <TT5>
 F:944-977/Domain: tetratricopeptide repeat homology <TT6>
 F:978-1010/Domain: tetratricopeptide repeat homology #status atypical <TT7>
 F:1011-1044/Domain: tetratricopeptide repeat homology <TT8>

Query Match 3.9%; Score 207.5; DB 2; Length 1119;
 Best Local Similarity 19.4%; Pred. No. 0.053;
 Matches 223; Conservative 184; Mismatches 309; Indels 341; Gaps 59;

```

QY 35 KQKOTOSASED -DMPEEDNERKTNSKENVSTV-----DETVSOLFSDGNS----- 78
Db 94 KDLQDSANQILKIKGQDNTKTNN--ENFDIAPNNRYIKDSTITTEYSRNDVGIDE 152
QY 79 -----NNSSKTESVSDPK-----QVPAKDEVT--OEASN 108
Db 153 DISEFKSKIPKIKPTNPKEDQIQSPNPKLSVNDQKLNLELTKKNLSGKNSSEN 212
QY 109 SSNDASKEVVPKODPSKKELETSTWEAVD---FYTRGDTLVGFSKSGINKLS----QT 161
Db 213 IINDSOKIENDKONTNLSKESKNSENILKTPDNSKYSNNNT-----TSIKKISSQKE 266
QY 162 SHVLVBHADG-----TOLTQVAPFAPDPKKTAIAVTSRLGNGKRSRLDIOKEI 215
Db 267 SELSPSPQTIIGKIYPRYSYLKKELEIIDDINTG---RVTLGKN---RL---KEL 314
QY 216 IDEGEIFNAVQ-----LTKLTIPINGYKSIIGODAFVNDKNIAYVN--254
Db 315 IKKG-LSNKQKQNNLEIENSKNKEASNLTLTLIKOIEPLNIPDPDY--KKEIQGLDK 371
QY 255 ---LPSLETTISDYAFHMSLKOVLPDN-----LKVIGELAPFDNOIGKLYLPHLI 305
Db 372 EDKKPQYLEDKSKVH---SIKPIDENTKSRQQAIDKLANE--FLKNV-----PND-A 418
QY 306 KLABRAFKSNRIOTVEFLGSKLVIGBASPODNNLARVWL.PDGLKIESRAFTGNPDEH 365
Db 419 QASHTTAQAANKIQHLEDKSKVHSIKPIDENTKSRQQAIDK---LNFELKNNPDAQ 473
QY 366 YNNQVVLRTTQGNPHQATENTYVND---DKSLWRATPDMDTKMLEDEFTYQKNSV 420
Db 474 ASHTLAQANKI-QHLEDKSKVHSIKPIDENTKSRQQAIDKLANE--EF-LKNPN 524
QY 421 TGSNKGLOVRNKNKLEIPIKONGITITIGNAPFNDPOSK--TIKTYDIEIK-LP 477
Db 525 DAQASKTLAQANKIOHLE-----DLKSKVHSIKPIDENTKSRQ 563
QY 478 STRIKGAFAPQSNLKSFEASEDLEIKGAFMNNRIGTL-DLKDLKIGIQAAPHNH 536
Db 564 QAIDKLANEFL--KNPNDAQASKTLAQAA-----NKIQHLEDKSK-----VHS 604
QY 537 IYAVL-----PESVOEIGRSAPFRONG--ALHMLFIQNKVKTIGENAFSLNKLESV 585
Db 605 IKPIDENTKSRQQAIDK--NEFLKNNPDAQASKTLAQANKIOHLEDKSKVHSIKPI 662
QY 586 NLSBQKQKTIIEVOAFEDNLSVVL--PNNLOQIR--EBAFKNHLEKVG-----633
Db 663 DLENTKSRQ---QAIDK--LNEFLKNNPDAQASKTLAQANKIOHLEDKSKVHSIKPI 716
QY 634 ---SSTLSQ---ITFNAFDONDGKRGFKGVVVRTNNSHMLADGERFIIDPKLSSTWV 687
Db 717 DLENTKSRQQAIDKLANEFLKNNPDAQASKTLAQAYANNNGDLKAENAYEKIILNTQIE 776
QY 688 D-----LEKVLKTIIEGLDYSTLRQTTQTOFREMTTAKAL--LSKSNLRQGEKOK 735
Db 777 DHVKGIIIRFKLKKYEHSISFP-QTIKLDPKK-KALHNKGIALMMLNKNKKAIESFEK 834
QY 736 FLOEAPFLGRVLDKALAKAEKALVTK-----ATKNHILERSINKAV-----780
Db 835 AIQ-----IDKNYGTAYYQKGIABEKNGDMQOAFASFNAJNLNKNPNYALAKAGIVS 886
QY 781 -----LAVNSAIKKAIVKRLLEKLDLTLJLVE-----808
Db 887 NMLGNFQSEBYLANFPAANAKKREIATVNLSTIKFENNLBESLETINKAIDANPKSE 946
QY 809 -----KGELAQATVQGVYLLKTPLPLEYYIIGLVNVPKSGKLIYALD-MSDT 857
Db 947 YLYLKASINKEKENYQNAISLYSLVIEKNP--ENTSAYINLAKAEKSGNKSQAISTLEKI 1005
QY 858 IIGSQDQAYGN-----PIINVEDNEGHTTLAVATLADVEGLYIK 897
Db 1006 INKNKALNANLILYKKEKNYQKALIEFERKAITN--SDIEAKNIATTLLEINDNTRAK 1063

```

```

QY 898 DILNSLSDKIKAIRQIPLAKYHRLGIFQAIRNAAAEADRLPTPKGYLNEVNYRKKOM 957
Db 1064 DLIR-BYTLKKNRNPAL---HALGIIETVENNNQDTLREL-----IKKPPNYKANE 1111
QY 958 ENKLKPV 964
Db 1112 --NIKKI 1116

RESULT 15
T28677
rhoptry protein - Plasmodium yoelii
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28677, C45521
R:Keen, J.; Sima, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A:title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
A:Reference number: Z20508, M0ID:95021522, PMID:7935623
A:Accession: T28677
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2269 <KEB>
A:Cross-references: EMBL:J27838; NID:9457145; PID:9457146; PIDN:AAA21304.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple
A:Reference number: A45521, M0ID:9110160; PMID:2270106
A:Accession: C45521
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 2131-2269 <KE2>
A:Cross-references: GB:M34283

Query Match 3.9%; Score 207; DB 2; Length 2269;
Best Local Similarity 18.8%; Pred. No. 0.16;
Matches 226; Conservative 206; Mismatches 432; Indels 340; Gaps 58;

QY 29 ERESVQKQOTOSASDDMEEDNERKTNSKENVSTV---ETVSOLFSDGNSNNSSKT 85
Db 989 DHDYNIKELSKHKQKNGYKTEADQNKKAIQKKELPEQYKEBYVTLNKTIVLKNKF 1048
QY 86 ESYVSDPKQVFK---AKPEVTOEAGNSNSDASKV-----EYVKQDTASKK----- 127
Db 1049 DKTNDKSKQIKKIKOAHNYCTLESQSKKKNMEIKNEKIHIDEVANNKSKKALTSIK 1108
QY 128 ---ETLETSTWEAKDVTYTRGDTLVGFSKSGINKLSQTSNHLVLPSSHAADGTLQVASFAP 184
Db 1109 VSEVPKTKIKIKINEIRKSD-----DLKKEITNDE-----KQISNL-- 1145
QY 185 TPDKTAIAEYTSRLGNGKPSRLDIOKEIIDEGRIFPNAVOLTKLTIPINGYKSIQDAF 244
Db 1146 ---SIDQETKYLENGK---QKLTLE--BLISLKKQKKNIE----- 1180
QY 245 VDNKNIAEVNLPSELTISDYAFHMSLKQVLPDMLKVIGELAFDNOIGKLYLERHL 304
Db 1181 -QKKEIDDEVN--SKIKNIETVYVQHKKNVEHIGIVEKINELAKTN--KNQLESFTELKPT 1235
QY 305 IKAERAFKSNRIQTV---EFLG---SKLVIGELASFQDNNLRVWL.PDGLKIESRAFT 358
Db 1236 IQHIIISFPANDLEGIDSDENLQKYNTEMGNIEEPIKSYNL-----ITVLETVSKESIT 1291
QY 359 GNPGEHYNQVVLARTTQGNPHQATENTYVNDPSLWRATPDMDTKMLEDEFTYQKN 418
Db 1292 -----YNOI-----ON-KRIDIQKELKNENIVNKAKEYLDYIK--ENEF--DR 1330
QY 419 SVTFNSKGLQAVRRNKNLEIPKOHNGI-----TTEIGD--NAFRYVDFQSKTL--- 466
Db 1331 IYTHFKKK-LNTVNDNPKNEYSKVNEGFNDISNITVKSSTJENSJLNTINQTKEMYAN 1389
QY 467 -----KTYDIEI-----KLPTIRKIGAFAPQSNLKSFEASEDLEIKGAFMNN 513
Db 1390 IVNNYYSYKYEAKENIFRNIPKLANTLN-----IKIKSSGIDLSKDI-KIALISYLD 1442

```


26	179	3.3	3418	2	US-08-603-753D-4	Sequence 4, Appl.
27	179	3.3	3418	3	ITS-09-099-753-4	Sequence 4, Appl.

278 PDNLKVI GELAFDNIGKLYLPRHLIKLAER-----AFKSNRIQTVEFLGSKLKI 33

```

QY 331 GEAFQDNNLRNWMVLPGLKEIESEAFNGPDENYNNVYLRTRGQNHOL----- 383
Db 1584 -EALIQLANQKS-----NEARTALO-NEHSNNGVAQASNGSNGHEHLYMPRAHK 1630
QY 384 -----ATENTYVNPDKSLMRATPD-----MDYTKMLEBDPTYQKNSVTSFSNKG 428
Db 1631 KSDAKQSIDNMY-NEQSTINTTTPDADTEBEQKALDKLK-IADAGVYNK-----V 1678
QY 429 QKVRANKKLEIPKQNGHITTEIGDNAFR-----NVDQSKLRKDYDEIK-----LPST 479
Db 1679 DQAOQNDQVDAKTEALDTITINIQANVAKKPSAEVLEDSK-----FEDLKQIQINATPNA 1732
QY 480 IRKIGAFAPQSNKISFEAS-----EDLBEI-----KEGAFM-- 511
Db 1733 TEEBKQDAILQRLNGRDEYVKLINODRRDNEVEQKXNIGLOELTIANPRKSDALQEL 1792
QY 512 -----NNRIGTLDLKD-----LIKIGDAFH-----INHI 537
Db 1793 QTKFISQTELLNNKKAIDNEEKDEKXRLLEISXKKTITINIAQOTNNQVDAKXONGMEI 1852
QY 538 YAIYLPESVQEIGSAPRQNGALHMF--GNKVKTIGMA----- 576
Db 1853 ATIIIPATIKTDATKAIIDKKAQOQVTLINGNNDATDEKKAERKLVEKAKIEASNTTNS 1912
QY 577 -----FLSNKLESV-NLSEQKQKTIETVQAFSDNLSSEV-----LPPMLQITREBAF 623
Db 1913 DTEBEVNGAKTNGLEKINNIPSTQYTNQAOEINDKQOBLIQUNTPDTEBEKQOAT 1972
QY 624 KR-----NHLKEVKGSSSTLSQITFNAFPQNDGDRFGKVVVRTHNSHM 668
Db 1973 NRVNAGLAQAIQININNAHSTQEVNESKTSIATIKSVOPNVIKKTATINSILOEANNQKT 2032
QY 669 LADGRFTIIDDKJSTWVDELEKVKIKIEGIDYSLRQTQOTQPREMTTACKALLSKSNL 728
Db 2033 LIGDGNATDDEKEKAKQLVTQKNEQIQKHESQDNQVNVKAAQAIATIKLINANNAHK 2092
QY 729 ROGQKQFLQEAQFAGVDDIKATAKAEKALVTCKATKNGHLLERSINKAV-----LA 782
Db 2093 RQ-----DAINITITLAEKSGDI--RANQDATTBEK--NTALQSIDDTFLA 2134
QY 783 YNNSAIKKANVYRLEKELDLTDLVEKGPPLAQATMVQGVYLLKTPLEPEVYIGLANYF 842
Db 2135 QARNNINGANTVAALYDE-----NLEDGKQKQOR-----TVL 2165
QY 843 DKSGLIYALMDSDTIGGQKQAYGNPLINVEDNE-----GYHTLAVATLADYE 892
Db 2166 STQYTTQAKADIAQAIQOQSTIIDQNAATTEKQEBALERLQNETNGVNDRIQALALANQ 2225
QY 893 GLYIKDIIINSLDKIKAIROLPLAKYHRLGIPQAIRNAAAEADRLLPKTPK----- 943
Db 2226 ---VDEKNNITLETIRNVEPIYIVKPKANEI--IRKKAABQTTLINQNOQDATTBEKQIA 2279
QY 944 -GYLNEVNPYRKKQNEKULKPYDYATPIFN-----KALPNEKIDGPRAKG 988
Db 2280 LGKLEEVNKQALNQVQASHSNNDVKTAEENGIKASEVHEPITIRKNAKQELQDQAOQI 2339
QY 989 HMINA---ETNN-----SVA-----VTPIRSEQQLKSSQDVN-----LPQTS 1023
Db 2340 DTINANKSTNEEKSALIDRVNVAKIDAINNITNATTQIVADAKNSGTSISQILPSTIA 2399
QY 1024 SKQNN 1027
Db 2400 VKTN 2403

```

RESULT 2
 US-09-134-001C-3159
 ; Sequence 3159, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; EPIDERMIDIS AND THERAPEUTICS

```

; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-3159

Query Match      4.1%; Score 222; DB 4; Length 10182;
Best Local Similarity 19.7%; Pred. No. 1.2e-05;
Matches 236; Conservative 200; Mismatches 435; Indels 324; Gaps 58;

QY 29 ERESVQEOQOASEDWFEDNERKTVSKENSTVDET-----SDLFSDGNSNNSSGKT 85
Db 8438 QREVMITNTNATTREKVAKNLDNAQ--LDKMETLQVVAHKNNITLNSKYLNEDSKY 8495
QY 86 ES-----VSDPKQVPPKAPPEVTOEASNSNDASKEVVPKODT-ASKK-----ETLESTW 135
Db 8496 QOQYDRVIADAQ-----LINQTNPTLEPYKVDIVKONVLANEKILFGAEKLSYDKS 8548
QY 136 EAKQFVTRGDTLVFSGSKGINKLSQTSILVLPASHADGTOQTQVASPAFTPDKKTAIAEY 195
Db 8549 NANDEIRHMYLNNAAQKQSLKDMT-----SHAARTVKQL----- 8584
QY 196 TSLRGENKSPRLDIDQKEIIDG---ELFNAYQLTKLTTPNGYKSIQOD----- 242
Db 8585 -----LQQAKLIDEMKSLJEDKQVVIDTTLPLN-YTEASEDKKQVQTV 8629
QY 243 ---AFPDNKIAVNP-----ESLETTSDVAFAMSLKQVYLPNKLAVIGELAFPDN 292
Db 8630 SHAQAIIDKJNGSNVSLDQVRQALEOLTQASENIDQDREAEAVHAN-QTIDQTLHNS 8668
QY 293 -QIGGLYLPBRLIKIAERAFKSNRIQVEFLSGKLKVGSEAFQDNNLRNWMVLPGL 350
Db 8669 LQOQTAKESV-KNATKLEELIATVSNNAQALN-----KWKQKLEQFINHADSVDNSNYR 8741
QY 351 KIESEAFQG-NPGEHNNQVYLRTRGQNE---HOLATENTYVNPDKSLMRATP-DM 403
Db 8742 QADDDKIIAYDEALEH---GQDIQKTNAQTNETKQALQOLIAETSLNGFEELINARPRAL 8799
QY 404 DYTEKLEBDFTYQKNSVTFGPNKGLQVRRANKLEIKQH---NGITTEI--GNAPRNV 459
Db 8800 EYISLEKINNAQKSALE-----DKVTQSHDL-LELEHIVNEGTMINDLMGELANAY 8851
QY 460 DFOGKTLAKYDLERIKLPSTIRKIGAFQSGNNL-KSPSESEDELEIKEGAFNN-NRIGT 517
Db 8852 NNNAFT-----KASINYINADMLRKDNTQALINNAKDALNTQOGQNLDPNADIT 8900
QY 518 LDLDKLIKIGDAFHINHIYAI-----VLPESVQEIGRSAP--RQNGALHMFITGNKYK 570
Db 8901 --FKODIFKTDALNGERLTAASKAKELIDSLKFNKAQFTHANDEIMWTNSIAQISR 8958
QY 571 TIGEMAFISNKLSEVNSIQKQKLTIEVQAS-----DNALSEVVLPEPNLOT 617
Db 8959 IVNQAFDINDMKS--LDELDNMQAFVQASNTYINSDEDLKQGFDAHLSNA-----RKV 9011
QY 618 IREBAFKNNHLKEVKGSSSTLSQITFNAFPQNDGDRFGK---XVVVRTHNSHMLADGR 674
Db 9012 LAKENGKULDEKQIQG--LKQVLEDTDALNGIORLSKAKAKAIQVQASLSY-INDAQR 9067
QY 675 FIID-----PKLSS-----TMDLEKVK-----LIEGIDYSLRQTQOTQ 711
Db 9068 HTAENNHNSDLSLANTLSKASDLDNAMKDLRTITNSGTSVNSVNYINADKYLQIE 9127
QY 712 FREMTTACKALSKS-----NLRQGEKQFLQEAQFAGVDD 750
Db 9128 FDEALQASATSKSISENPATIEVYLGISQAIYDTKNALNGB-QALATEKSKDKLKLKGL 9186

```


APPLICANT: Hatley, Calvin
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: No. 6093809e1 Telomerase
 NUMBER OF SEQUENCES: 225
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/851,843A
 FILING DATE: 06-MAY-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002930US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 872 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 US-08-851-843A-8

Query Match 3.7%; Score 200; DB 3; Length 872;
 Best Local Similarity 20.3%; Pred. No. 1.2e-05;
 Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

QY 14 TTVSVVTVYSGEVYGL-REESYK-SEQTQASDEDDWFEENENK-55
 DB 17 TWLDFVLQNLVYKQIIEHYKTQQOQIKEDDKLTKFKQDQDQSGNDDDEENNSNQ-76
 QY 56 -TNVSKENSTVDETVSDLPSPDGSNNSSKTEVSDPQVPAKXDEVQASN-108
 DB 77 QELLRRVNOIKQOVQILKVKVSKVEKDLN-EDEN-111
 QY 109 SSNDASKVVPKODTASKEETETSTWEA-KDFTVSGDTLVGSKSG-154
 DB 112 KNGGSGEQV-KERQRTITEGVKQNLVFNMDYGLDNLNESCGRHRRRETDT-164
 QY 155 -INKSQTSHVLPBHADGTQTVASPAFTPDKTAIAETSRILGNGKSRDIDOK-213
 DB 165 DTEKWEFISH-DQKNVSIYA-NQTSYCMWLKOVFNKNYDHLNVSIN-211
 QY 214 EIIDGEIF-NAYQLTLPNGYKSGIDAFVNNK-NIAEV-253
 DB 212 RIETAEFAFDPSFTIKLT-NSYQYINIDVFNENNLCTALRLFLSLERFNILNR-270
 QY 254 -NLPESLETTSDYAFAMSLKQVCLPDNLKVIETLAFPNQIGK-297

DB 271 SSTRNQVNPKEKIGLELTTPAVFESHRLQIHLOVPEAFQYLVNSSQSISTVKDSOQ-330
 QY 298 LY-LPRHILKAENAFNSNRQVTEPFLGSKYIGEMS-334
 DB 331 VYFSTDLKLVDTNKKVQDYEFKPLQEPFR-LTHVSQQAIPVATNAVENLVILKKVKA-388
 QY 335 FQNNNLRNWMLPGLKEISEAFVGNPDGHYNNQVLPRTGONPHOLATENTYVNDK-394
 DB 389 -NMLNLSIP-TQNPDPFYFNLOHILKEFGLERINILTKQKL-E-428
 QY 395 SLWRAPPDMDYTKLEEDF-TV-QKNSVTGFSNKG-LQKVRNKNLEIPKQHNITIT-449
 DB 429 NILLSIKQSNKMLKFLNLFYTYAQSFRQILKQATTKINLKNKNGQSEETPEKQETPS-488
 QY 450 EIDGNAPRNDPQSKTKRKVDLEIKLPST-IRKIGARAFQSNKSKPEAS-EDLEE-504
 DB 489 E-STGCKRFPDHSLELTEDF-SVNLQATQEIYDSHLKLLIRSTNLIKFKLSYKYMEX-546
 QY 505 IKGAF-MNNRIGTL-DLKOKLIKIGDAFHNHLYAIVLPESV-QET-G-550
 DB 547 SKMDTFIDLNKITYETLNNLKRCSVNISNP-HGNISYELNNKOSTFYKPKLTLLNQELQA-604
 QY 551 RSAFRONGALHMFIGNKVTIGEMAFLSNKLBSV-NLSQOKLTREVOAFS-602
 DB 605 KYTEKON-EFOFNNVKS-AKISSLSLEIDIDSLCKSIASCKNLQNV-650
 QY 603 DMLSEVVLPPNLOTIEEAFKRNHL-KEVGSSTLSQITENA-PDONDGDRFGKV-658
 DB 651 -NITASLLYPNNID-KNPKNKPLFPKQPEOLKNLBNVINCILDQ-695
 QY 659 VVRTHNSHMLADGERFIDPDKLSTWVLEKVLKIEGLDVSYTBQTTQ-709
 DB 696 -HINLSISELENNKKIKAPLKRYYLQY-LDYKLFKTLQOLBELNQVY-745
 QY 710 -TOPREMTTAKKLLSKNSINROGEK-FLQEQFLGKVDIDKALA-KA-756
 DB 746 INQOLEELVSE-VHKQWENHKQAFEPYELCEPIKSSQTLQIDPQNTVSDSI-801
 QY 757 EKALVTKKATNGHILRSIRINKAVIAYNSAIKKANYKLEKEDLLTDLVGGKPLAQ-816
 DB 802 KKLIESISESKTHHYLRNPSQ-SSSLIKEN-EELQELKACDEKGVYKA-851
 QY 817 TMQGVYLLKTPLPPEYIIGLVNYPD-843
 DB 852 -YKFPPLCP-TGYTD-866

RESULT 5
 US-08-851-843A-54
 Sequence 54, Application US/08851843A
 Patent No. 6093809
 GENERAL INFORMATION:
 APPLICANT: Cecchi, Thomas R.
 APPLICANT: Lingner, Joachim
 APPLICANT: Nakamura, Toru
 APPLICANT: Chapman, Karen B.
 APPLICANT: Morin, Gregg B.
 APPLICANT: Hatley, Calvin
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: No. 6093809e1 Telomerase
 NUMBER OF SEQUENCES: 225
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/851,843A
 FILING DATE: 06-MAY-1997
 CLASSIFICATION:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 CLASSIFICATION:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 CLASSIFICATION:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002930US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 54:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 872 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: peptide
 US-08-851-843A-54

Query Match 3.7%; Score 200; DB 3; Length 872;
 Best Local Similarity 20.3%; Pred. No. 1.2e-05;
 Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

14 TTVSVTVYSGEVYGL-----REESVY-----OEQOSASEDMFEDMERK- 55
 17 TNDLVONLEVYKSGQHEHYKTOOQKEEDLKLFKNDQNGSNDDEENSNKQ 76
 56 -----TWVSKENSTVDETSDLPFSDGNSNNSSKTESVSDPKQVPKAKPEVTOEASN 108
 77 QELLRRVNOIKQOVOLIKKVKSGVKEKDLN-----EDEN 111
 109 SSNDASKEVYPKODTASKEETLETSTWEA---KDFYTRGDTLVGFSKSG----- 154
 112 KKNGLSEHQV-----KEQQLRTITEEQVYKYNLVFNMDYQDLNESGGHRRHRETTY 164
 155 -INKLSQTSHLVPSHADGTOLTVASFAFTPDKTAIAEYTSRLGENGKPSRLDIDQK 213
 165 DTEKWEPEISH-----DQKNYISYIA---NOKTISYCWMLKDYINKNKNYHILNVSTIN 211
 214 EIIDEGEIF---NAYQTLTKLTIPNGYKSGIGQDAFVDNK-----NIAEV- 253
 212 RLFTAEAFYAPDDPSQITKLT--NNSYQTVNIDVFNNDLCLALLRFLSLERFNILNIR 270
 254 -----NLPESLETISDYAFAMSLKQYKLPENLAVIGELAFPDQIGK----- 297
 271 SSYTRNQYNFEKIGELLETTPAVVFSRHLOGHLOVPCAFQYLVNASSQISVKSQLO 330
 298 LV-----LPRHLIKLAERAFKSNRIQTVFPLSKLKVIGEAS 334
 331 VYESTSTDLKVDNKKQDYKFLQEPFR--LTHVSQQLIPVSATNAVAVNLVLLKKVYHA- 368
 335 PQDNNLRNVMPLDGLKEISEBAFTGNPDDEHYNNQVLRFRFGQPHQALATENTYVNDK 394
 389 -----NMLVSIIP-----TOFNDFEFVNLQHLKLEFGLEPILTKQKQ-----E 428
 395 SLWRATPDMQYTKWLEEDF--TY--OKNSVYGFSSKSG--LQKVRNKKLETIPKONGITIT 449
 429 NLLISTIKQSTKNLFLRLNFYTYVAQETSRKQIILKQATTTIKVLLKNNKNOESTPEYKDETPS 488

450 EIGDNAFRVNDPSQTKLRKYDLERIKLPST---IRKIGAFAPQSNLKSFEAS--EDLEE 504
 488 E-STGKMFEPDHLSELTELEDF-SYNLQATQETIYSIKLRLSRNLLKPKLSYKEMEK 546
 505 IKEGAF--MNNRIGTL-DLKDLIKIGDAFHHIYAVLPESV-----OEI--G 550
 547 SKMDTFIDKNIYETLNNLKRCSVNISNP--HGNISYELTNDSTFFYKFKLTLNQLQHA 604
 551 RSAFRONGALHLMFGNKYKTIGEMAFISNKLSEV-----NLSQQLKTIYQAS 602
 605 KTFPKQ-----EFQFNWKS---AKIESLSLESDIDLSLCKSIASCNLQNV----- 650
 603 DNALSERVLPVNLQITREBAFGRNHL---KEYAGSSTISQITFNA-FDONDDKRRGKY 658
 651 -NIMSLEYNNIQ--KNPFNKPNLFPKQFOLKNLENVINCLDQ----- 695
 659 VVRTHNNSHMLADGERFIIDPKLSTWVLEKVLKIIIGLDYSTLRQTTQ----- 709
 696 -----HILNSISFLEKNKKIKAFILKRYLLQY--LDYTKLFKTLQQLPELNQY 745
 710 --TORREMTAGKALSKSNLQGEKQ-----FLOBAQFPLGRVLDKALA---KA 756
 746 INQOLEELTVSE---VHQVWENHKQAFYEPCEFIKESQTLQDIDPDQTVSDSI 801
 757 EKALVTYKTKKQGHLLERSINKAVLAVNNSAIKKANVKKLELDLTLVGEKQPLAQ 816
 802 KILLESISESKYHHYKRLNPSQ-----SSSLKSEN-----BEIGBLACCEKGVLYVA 851
 817 TWQGVYLLIKTPLPLPEYITGLNLYFD 843
 852 -----YKFPPLCLP-----TGYTVD 866

RESULT 6
 US-08-974-549A-221
 Sequence 221, Application US/08974549A
 Patent No. 6166178

GENERAL INFORMATION:

APPLICANT: Czech, Thomas R.
 APPLICANT: Lingner, Joachim
 APPLICANT: Nakamura, Toru
 APPLICANT: Chapman, Karen B.
 APPLICANT: Morlin, Gregg B.
 APPLICANT: Harley, Calvin B.
 APPLICANT: Andrews, William H.
 TITLE OR INVENTION: Human Telomerase Catalytic Subunit
 NUMBER OF SEQUENCES: 727
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/974,549A
 FILING DATE: 19-NOV-1997
 CLASSIFICATION: 536
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/851,843
 FILING DATE: 06-MAY-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/854,050
 FILING DATE: 09-MAY-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/911,312
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/912,951
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/915,503
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US97/17618
 FILING DATE: 01-OCT-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US97/17885
 FILING DATE: 01-OCT-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph Ted
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002610US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 221:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 872 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-974-549A-221

Query Match 3.7%; Score 200; DB 3; Length 872;
 Best Local Similarity 20.3%; Pred. No. 1.2e-05;
 Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

14 TTASVVTYSQEVYGL-----REESVK-----QOTQASASDDMFEEDEBERK- 55
 17 TNLDFVLQNLVYYSQIHYKTCQQQIKKEEDLKLFKNQDQDNGSDDDDERNNNSKQ 76
 56 -----TNVSKENSTVDETVSDLESDNSNNSSKTESVSDPKQVAKAPEVTOEASN 108
 77 QELLRRVNOIKQOVQOLIKKVGSKVEKDLNLN-----EDEN 111
 109 SSNDASKVEVPKQDTASKKETLETSTWEA---KDFVTRGDTLVGFSKSG----- 154
 112 KKNGLSEGOV-----KEBQIRITTEQOVKQNLVFMMDYQDLNBSGGHRRHRETDY 164
 155 -INKLSQTSHLVPSHADGQTLTVASFAFTPPDKKIALAETSLGSKSPSLDIDQK 213
 165 DTEKFEIISH-----DQKNVYSIYA---NOKTSCWMLKDYFNKNVYDHLNVAISN 211
 214 EIDGGEIF--NAVQTLKTLTPNGYKSGIDAFVDNK-----NIAEV- 253
 212 RLLETAETFAFPDESQITKLT-NSYQTVNIDVNPNDNLCTIALRLPILSLERFIILNIR 270
 254 -----NLPSLETTISDYAPAHMSLKQVKLPDLKVIIGELAFDNOIGK- 297
 271 SSYTRNQVNEFEKIGLELFTIFAVVSHRLQGIHLQVCEAPQYLVNSSQISVDSOLO 330
 298 LY-----LPRHLITLARARAKSNRIQVPEVFGSLKVIIGAS 334
 331 VYSFSTDKLVDTNKVQDYEFKLOEPPR-LTHVSOQALPVSATNAVENNLVLLKVKKAH- 388
 335 FQDNMLRNVMLPDGKLEKLESEAFTPGNGDEHANNQVLTFRGQNPQOLATENTVNDPK 394
 389 -----NLNIVSIP-----TQENPFYFVNIQHLKLEGLEENILITQKLT-----E 428
 395 SLMRATPPMDYTKWLEDF-TY--QKNSVTGSSNKG--LQKVRANKNLEIPQHGNGITTT 449

429 NLLLSIKQSKNLKFLNLFYVVAQETSHKQILKQATTIKLNKNNQNEETPDKDTPS 488
 450 EIGDNAFRVNDVQSKTKRKVDLEITKLPST---IRKIGAFAPQSNLKSFEAS--EDLE 504
 489 E-STSGMKFPDHLSELTELEDF-SVNLQATQETIYDSLHKLLIRSTNLKPKFKUSKYEMEK 546
 505 IREKAE--MNNRIGTL-DKKDLIKIGDAAFHNIHIAVLPEVS-----QRI-G 550
 547 SKMDTIDILKNITETIANLKRCGVNISNP--HONISYELTNKDSOTYKFKTLINQELQHA 604
 551 RSAFRONGALHLMFIGNKVKTTIGEMAFISNKLSEV-----NISEQQLTEVQARS 602
 605 KYTFKQN-----EFQNNNVYS--AKIBESSLESLDIDSLCKSIASCNKLQNV----- 650
 603 DNALSEVLPRLNQTIRREAFKKNHL--KEVGSSTLSQITFNA-FPDNDGKRGKKV 658
 651 -NIIASLILYFNNTQ--KQFENPNLILFFKQFQKLNLENVSNICILDQ----- 655
 659 VVTRHNNSHMLADGERFIIDPDKLSTWVLEKVKIIEGLDYSTLRQTTQ----- 709
 696 -----HILNSISEFLEKNKKIKAFILKRYLLQY--LDYTKLFKTLQQLPELNQY 745
 710 --TQFEMTTAGKALLSKSNLRQGEKQK-----FLQAPQFLGRVIDDKAIA--KA 756
 746 INQQLBELTVSE-----VHKQWENHOKKFEYEPICEFIKESQTLQIDPDQNTVSDSI 801
 757 EKALVTRKATKRNKHLERSINKAVLAVNNSAIKAVKRLKEKEDLITLVEGKPLAQA 816
 802 KKLTESTSESKYHHYLRNLSQ-----SSSLINSEN-----EIQELKKACDKGVLYKA 851
 817 TMOGVYLLKTPPLPEYVIGLVNYPD 843
 852 -----YKFPPLCP-----TGYYD 866

RESULT 7
 US-08-854-050-8
 Sequence 8, Application US/08854050
 Patent No. 6261836
 GENERAL INFORMATION:
 APPLICANT: Cecchi, Thomas R.
 APPLICANT: Lingner, Joachim
 APPLICANT: Nakamura, Toru
 APPLICANT: Chapman, Karen B.
 APPLICANT: Morin, Gregg B.
 APPLICANT: Harley, Calvin
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: No. 6261836el Telomerase
 NUMBER OF SEQUENCES: 225
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/854,050
 FILING DATE: 09-MAY-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/851,843
 FILING DATE: 06-MAY-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 536
 FILING DATE: 25-APR-1997
 CLASSIFICATION: 536

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002300US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 872 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 US-08-854-050-8

Query Match 3.7%; Score 200; DB 3; Length 872;
 Best Local Similarity 20.3%; Pred. No. 1.2e-05;
 Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

QY 14 TVSVVTVYSQEVYGL-----REESVK-----OEOQSASEDDWFEEDNERK- 55
 DB 17 TNDPVLQNLVYVYKQEHYKTOQQQKKEEDLKILKFKNDQDSNSNDDEDENNNKQ 76
 QY 56 -----TWSKENSIVDEVYSDLPFGNSNNSKTESVSPDKQVPAKPEVTOASN 108
 DB 77 QELLRVNOIKQOVQILKVKSKVKXDILN-----EDEN 111
 QY 109 SSNDASKEVVKQDTASKETLETSTWEA---KQVTRGDTLVGFSKSG----- 154
 DB 112 KKGGLSEQY-----KEQLRTITEQYKQVLVNMDDYQDLNLSGGHRRHRTDY 164
 QY 155 -INKLSQSHLVLPASHADGTQTLQVAFAPTPDKTAIAEYTSRLGNGKPSRLDQK 213
 DB 165 DTEKWFESH-----DQKNYVSIYA---NQKTSYCMWLKDYVKNKNYDHLANSIN 211
 QY 214 EIIDEGEIF---NAYQITKLTIPNGYKSGODAFVNDK-----NLEV- 253
 DB 212 RLTEAEFYAFDDFSQITKLT--NNSYQTVNIDVFNMLCILALRLSLERENILIR 270
 QY 254 -----NLPESETISDVAFAMSLKQVLPNLKVIQELAFEDNOIGK----- 297
 DB 271 SSTTRNQYNPEKIGELLETIFAVVFSRHLOGHIQVPCAFQYLVNSSSQISVDSOLQ 330
 QY 298 LY-----LPRHLIKLAERAFKSNRLQTVFELGSKLKVIGEAS 334
 DB 331 VYFSSTDLKLVDTNKVQDYFELQEPFR--LTHVSGQALIPVATVAVENLVLLKVKYGA- 388
 QY 335 FQDNNLNFWLPLDGLKIESEAFVGNPDGHYNNQVLRIRGTGNPHQALATENTYVADK 394
 DB 389 -----NMLVSIIP-----TQFNPDYFVNLQHLKLEFGLEPILTKQK-----E 428
 QY 395 SLWRATPDMQYTKLEEDF--TY--QKNSVTFGFSKSG--LQKVRNKLKLEIPKQNGITIT 449
 DB 429 NMLISIKQSKNKLKRLANFTTYVAQETSRKQILKQATITIKKNNKQOESTPEKDETPS 488
 QY 450 EIGDNARVNDPQSKTLRYDLEIKLPST--IRKIGAPAFQSGNNLKSFAS--EDLEE 504
 DB 489 E-STSGMKFPDHLSELTELEDF--SVNQAQOEIYDSLKLILRSTNLKKEFLSYKEMEK 546
 QY 505 IKGEAF--NNNRIGTL--DLKDKLIKIGDAFHNIHAIYVPEV-----QEL--G 550
 DB 547 SKMDTFIDLKNIYETLNNLRCSVNISNP--HGMSIYELTNKSTPYKFKLITNLOEIOHA 604
 QY 551 RSAFRQNGALHLMFIGNKVKYTIQGMATLSNKLSEV-----NLSEKQULKTTEVQAFS 602

DB 605 KTFKQK-----EFQNNVYS---AKTESSELSLEIDSLCKSINCKNLQNV----- 650
 QY 603 DNALSEVYLPNLOTTRERAEAFKRNHL---KEVGSSTLSQITPNA--PDQNDGRRFQKV 658
 DB 651 -NIIASLYVNNIQ---KNPFKNPILFFQPFQNLNLEVSINCLIDQ----- 695
 QY 659 VTRHNNSHMLADGERPIIDPKLSTWDLKLVKIIIEGLDYSTRQTO----- 709
 DB 696 -----HINISSEFLEKNNKIKAFILKRYVLYQY--LDYTKLETKLQQLPELNQVY 745
 QY 710 --TQPEMTTAGALSKSLRQGEKQK-----FLQRAQFFLGAVDLDKALA---KA 756
 DB 746 INQLEELVSE---VHKQWENHOKAFEPYLCETIKSSQTLQIDPDQNTVSDSTI 801
 QY 757 EKALVTKATKNGHLLERSINKAVLVNNSAIKKAVKRLKEKEDLLTDLVEGKPLAQ 816
 DB 802 KKLIESISESKYHHYRLNPSQ-----SSSLIKSEN-----ERIQELKXKDEKGVLYKA 851
 QY 817 TMQGVYLLKTPPLPBYIYGLNVYFD 843
 DB 852 -----YKFPPLCLP-----TGTYYD 866

RESULT 8
 US-08-854-050-54
 Sequence 54, Application US/08854050
 Patent No. 6261836
 GENERAL INFORMATION:
 APPLICANT: Cech, Thomas R.
 APPLICANT: Lingner, Joachim
 APPLICANT: Nakamura, Toru
 APPLICANT: Chapman, Karen B.
 APPLICANT: Morlin, Gregg B.
 APPLICANT: Harley, Calvin
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: No. 6261836el Telomerase
 NUMBER OF SEQUENCES: 225
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/854,050
 FILING DATE: 09-MAY-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/851,843
 FILING DATE: 06-MAY-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002930US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 54:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 872 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: peptide
 US-08-854-050-54

Query Match 3.7%; Score 200; DB 3; Length 872;
 Best Local Similarity 20.3%; Pred. No. 1.2e-05;

Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

14 TTAVVVTYSQEVYGLL-----REESVK-----QEQTOSASEDMFEEDNERK 55
 17 TNLDPVTLNLEVYSQIEHYKTKQEQQIKEDLKLKFNQDQDGSNGDDDENNSKQ 76
 56 -----TNVSKNSTVDETVSDLFSDGNSNNSSKTESVVSDEKQVKAPEVTQASN 108
 77 QELLRRVNOIKQVQLIKKVGSKVEKDLNLTN-----EDEN 111
 109 SSNDASKVEVPKODPASKKELETSTWEA---KDFYTGDLVYGFSGSG----- 154
 112 KKNGLSBOOV-----KEQQRITTEQVYKQNVFNMDYQOLDINESGHRHRRFTDY 164
 155 -INKLSQTSHLVLPASHADGTOLQVASFAPTPDKKIAIAYTSRLGNGKPSRLDIDQK 213
 165 DTEKMFELSH-----DQKNVSIYA---NOKTSYCWMLKQYFNKNYDHLNVAIN 211
 214 EIIDGEELF---NAVQTLKLTIPNGYKSIQGDAPVDNK-----NIAEV- 253
 212 RLETFARFYAPDFEQTKLT---NNSYQTVNIDVNFNNMLCILALRFLSLERFLLNIR 270
 254 -----NLPSLETISDYAFAMSLKQVRLPDLKVIYIGELAFDNOIGK----- 297
 271 SSYTNGQNVFEKIGLELTIIPAVVSHRLQGIHQVCEAFQYIVNSSQISVDSQLO 330
 298 LY-----LPRHLTKLAERAFKSNRIQTVFELGSLKVIYIGAS 334
 331 VYSFSTDLKLVDTNKVQDFKFLQEFPR-LTHVSGQALPVSAKTNVEMNLVLLKKVKA- 388
 335 FQDNLNRVNLDPGLEKISEAFIGNPGBEHNNQVLRTRGQNPQOLAHENTYVNDK 394
 389 ---NLNVSLP-----TQFNDFYFVNLQHLKLERGLEPNILTKQKL-----E 428
 395 SLMRATPMDVTKMLEDF-TY--OKNSVTGFSNKG--LOKVRKKNLEIIPKOHNGITIT 449
 429 NLLLSIKOSKULKFLNLFYVVAQETSRSQILKQATITIKLNKKNKQDEPTEPDETPS 488
 450 EIGDNAFRVNDPQSKTLRKYLEIKLPST--IRKIGAFAPQSNLKSFEAS--EDLEE 504
 489 E-STSGMKFPHLSHLELEDF-SVNLQVTOEYDLSHLTLIRSTNLKKFKLSYXEMEK 546
 505 IYKGAFL--MNNRIGTL-DLKDKLITIGDAAPHINNTIYAVLPESV-----OEI--G 550
 547 SKMDTFIDLNKIYETLNLKRCVSINSP--HGNISYELTNKDSFFYFKLTNLNDELHA 604
 551 RSAPFRNGALHMFIGNKVTIGENAFISNKLSEV-----NLSEQKOLKTIIVQAFS 602
 605 KTFYKQN-----EFQPNVKS---AKIBESSLSLEDDIDSLCKSLASCKNLQNV----- 650
 603 DNALSEVVLPPNLQITREAFKRNHL---KEYKSGSTLSQITFNA-FQNDGDKRFGKGV 658
 651 -NIASLAPYNNIQ--KMPFKPNILFPKQFQKJNLENVINCILQD----- 695
 659 VRTTNNSGMLADGRFIIIDPKLSTWVDELEKVIKIEGLDYSYTLRQTO----- 709
 696 -----HILNSISEPFEKKNKIKAFILIKRYLLQYV--LDYTLKFLTLQQLPELNOVY 745

QY 710 --TOPREMTAGKALLSKSNLRQGEKOK-----FLOEAPFGLGRVDIDKAI---KA 756
 DB 746 INQLEELTVS-----VHKQVWENHKQKAYEPLECEIKKSSQTLQIDPDQVTVSDSI 801
 QY 757 EKALVTKATKNGHLLERSINKAVLYNNSAIKKAVKRLKEKELDLITDIVEGKPLAQ 816
 DB 802 KKLIESISESKYHVLRLNPSQ-----SSSLISEN-----EFIQELKKDEKGVLYKA 851
 QY 817 TMQGVYLLTPELPPEYTYIGLVYFP 843
 DB 852 -----YXKFPCLP-----TGTYD 866

RESULT 9

US-09-430-323-8
 Sequence 8, Application US/09430323

Patent No. 6309867

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.

Lingner, Joachim

Nakamura, Toru

Chapman, Karen B.

Morin, Gregg B.

Harley, Calvin

Andrews, William H.

TITLE OF INVENTION: No. 6309867el Telomerase

NUMBER OF SEQUENCES: 225

CORRESPONDENCE ADDRESS:

ADDRESS: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/430,323

FILING DATE: 29-Oct-1999

CLASSIFICATION: <Unknown>

Prior APPLICATION DATA:

APPLICATION NUMBER: US 08/854,050

FILING DATE: 09-MAY-1997

APPLICATION NUMBER: US 08/851,843

FILING DATE: 06-MAY-1997

APPLICATION NUMBER: US 08/846,017

FILING DATE: 25-APR-1997

APPLICATION NUMBER: US 08/844,419

FILING DATE: 18-APR-1997

APPLICATION NUMBER: US 08/724,643

FILING DATE: 01-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002930US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 872 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 8:
 US-09-430-323-8

Query Match 3.7%; Score 200; DB 4; Length 872;
 Best Local Similarity 20.3%; Pred. No. 1.2e-05;


```

Db      165 DTEKFEFISH-----DQKNYSIYA---NQKTSYCWMLKQYFNKNNDHLNVSIN 211
Qy      214 EIDEGEIF--NAVQTLKLTIPNGYKSIIGODAFVDNK-----NIAEV- 253
Db      212 RIETAEFAFADDSQITIKLT--NNSYQVNIIVNPFNNLCIALALRFLSLERFILLINIR 270
Qy      254 -----NLPSLETISDYAFAMSLKQVKLPDNLKVIIGELAFPDNOIGK----- 297
Db      271 SSYTRNQVNFKEKIGELLETFPAVNVSHRHLQGIHLOVPEAOQYLVNSSQISVQDSQIQ 330
Qy      298 IY-----LPRHLIKLAERAFKSNRIQTVFPGSKIKVIGENS 334
Db      331 VYSFSTDLKVDTNKVQDYEFKFLQEFPR-LTHVSQQAIPVSATNVAENLVNLVKKVKA- 388
Qy      335 FODNNLRNVMLPDGEKISEAFATGNGPEEHNNQVLFTRGQNPQOLATENTYVNPDK 394
Db      389 ----NINLVISIP-----TQNFDFYFVNLOHLERFLEPNILTKQKL-----E 428
Qy      395 SLMRATPDMDYTKMLEEDF-TY--QKNSTYGFNSNGK--LOKVRNNKMLEIPKQHNQITIT 449
Db      429 NLLLSIKQSKMLKFLRLNRYTVAQETSQKQILKQATTTKLNKNNQOETETEDTETS 488
Qy      450 EIGDNAFERNVDFQSKTLRKYLDEIKLPST--IRKIGAFAPQSNLKSFEAS--EDLEE 504
Db      489 E-STGGMKPFHDLSLTELDEF-SVNLQATQEIYDSIHLRLIRSTNLKFKKLSYKEMEK 546
Qy      505 IRECAF--MNNRIGTL-DLKDKLIKIGDAAPHINHYAIVLPESV-----OEI-G 550
Db      547 SMDJTFIDLNKTYETLNKLRCSVNSNP--HGNISYELTNKSTFFYFKLTINQLOLHA 604
Qy      551 RSAPRONGALHLMFIGNKYKTIGEMAFLSNKLSEV-----NLSEKQKLTIVQAFS 602
Db      605 KTFEQN-----EFCQNNVKS--AKIESSESLESDIDSLCKSLASCNLQNV----- 650
Qy      603 DNALSBEVLPPNLQITIREAFKRNHL--KEVKGSSSTLSQITFNA-FQDNDGKRGKKV 658
Db      651 -NIIASILYPNNIQ--KAPFNKPNLLEFKQFEQLKLENVNSINCILDQ----- 695
Qy      659 VRTNNNSHMLADGERFIIDPKLSSTWDLKYLKIHLGIDYSLRQTO----- 709
Db      696 -----HILNISISEFLEKKNKIKAFILIRKRYLLQYV--LDYTKLFKTLQOLPELNOY 745
Qy      710 --TOPREMTAGKALLSKSNLRQGEKOK-----FLOEAOFFLGRVLDYALA--KA 756
Db      746 INQJLEELTVSE-----VHQVWENKQKAFYEPLCEFISSQIQLIDFDQNTVSDSI 801
Qy      757 EKALVTKKATKNGHLEERSINKAVLAVNNSAIIKANVRLKLELIDLDLVGKGPLAOA 816
Db      802 KILLESISESKYHLYRLNPSQ-----SSSLIKSEN--EIOELLKACDEKGVLYKA 851
Qy      817 TWWGVVYLKTPLEPEYIIGLVNVPD 843
Db      852 -----YKFPCLCP-----TGYIID 866

```

RESULT 11

US-09-402-181B-221
Sequence 221, Application US/09402181B
Patent No. 6610839

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.
Langner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin B.
Andrews, William B.

TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 633
CORRESPONDENCE ADDRESS:

ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor

```

CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/402,181B
FILING DATE: 29-Sep-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Ausubus, Scott L.
REGISTRATION NUMBER: 42,271
REFERENCE/DOCKET NUMBER: 015389-002620US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 221:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 221:
US-09-402-181B-221

Query Match      3.7%; Score 200; DB 4; Length 872;
Beet Local Similarity 20.3%; Pred. No. 1.2e-05;
Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

Qy      14 TTVSVYVYSQVYGLD-----REESVK-----QROTQASQEDDMEEDNERK- 55
Db      17 TNLDFVLQNLLEVYYSQIIEHYKTOOQOIKEEDLKILKPKNQODQDGSNDDEENNSKQ 76
Qy      56 -----TNVSKENSTVDVETVSDLESQNSNNSSSKTESVSDPQVYKAKPEVYQESN 108
Db      77 QELLRNVQIKQOQOLIKKVGSKVEKDNLN-----EDEN 111
Qy      109 SSNDASKVVPKODTASKEETLETSTWEA--KQFVTRGDTIVGFSKSG----- 154
Db      112 KKNGLSEQV-----KEQQLRTITEQVQKYNLVFPMQDQLDLNBSGGRRRRRETDY 164
Qy      155 -INKLSQSHLVLSHAADGTQLTOVASFATTPDKKTALIAVTSRLGENKPSRLDIDQK 213
Db      165 DTEKFEFISH-----DQKNYSIYA---NQKTSYCWMLKQYFNKNNDHLNVSIN 211
Qy      214 EIDEGEIF--NAVQTLKLTIPNGYKSIIGODAFVDNK-----NIAEV- 253
Db      212 RIETAEFAFADDSQITIKLT--NNSYQVNIIVNPFNNLCIALALRFLSLERFILLINIR 270
Qy      254 -----NLPSLETISDYAFAMSLKQVKLPDNLKVIIGELAFPDNOIGK----- 297

```

```

Db 271 SSYTRNQVNEFKIGELLETFAVVFESHRLQGHVPCFAFQVLANVSSQISVXSQLO 330
Qy 298 LY-----LPRHLIKLAERFAFKSNRIQTVFELSKKVIQGEAS 334
Db 331 VYSESTDKLVDNKNQVDFKLOEFPRLTHVSOQAIPIVATVAVNANVLAKKVGHA- 388
Qy 335 FQNNLEHNVLPDGLKIESEAFNGDEHYNNQVVLRTGTQNPQHLATENTYVNDK 394
Db 389 ----NNTVSIIP-----TQPFDFYFVNLQHLKLEFEPNITLKOKL-----E 428
Qy 395 SLWRAPDMDYTKMLEDF-TY-QNNSVTFGSKNG-LQKVRNNKLLEIPKQNGITIT 449
Db 429 NLILSIKSKNKLKFLRNFYTYVAQETSRKQILQAVTITIKLNKNNKQGEFEPKDETPS 488
Qy 450 EIDGNAPRNVDFQSKTKRKTDLBEIKLPST--IRKIGAFAPQSNLKSFEAS--EDLEE 504
Db 489 E-STGSKFPHDSELTLEBDF-SVNLQATQEIYDSLHKLIRSTNLKKEFLSKYKEMEX 546
Qy 505 IKGAF--MNNRIGTL-DLKDOKLIKIDAAFHINHYAIVLPESV-----QET--G 550
Db 547 SKQDTFIDLNKIYETLNLKRCGSNINSP--HGNISETLNKOSTFYKFKLTINQELQHA 604
Qy 551 RSAFRONGALHMFIGNKRVKTIGEMAFLSNKLESV-----NLSEKQOLKTIYQAFS 602
Db 605 KYTFKQK---EFOFNKVS---AKIESSLESLEDIDSLCKSIASCKLQNV----- 650
Qy 603 DNALSEVLPNNIQTIREAFKRNHL--KEVKGSSLTLSQITTRA-PDQNDGDKRFGKKV 658
Db 651 -NIIASLTYNNQ--KNPFKNPMLFFKQFQOLKXLENSINCILQ----- 695
Qy 659 VVRTHNNSHMLADGERFIIDPKLSTMTVDLEKVLKIEGLDYSTLRQTLQ----- 709
Db 696 ----HILNISESELEKNNKIKAFILKRYLLQY--LDYTKLFKTLQOPELANQY 745
Qy 710 --TQFRMTTAGALISKSNLRQGEKQK-----FLQAPQFELGVDLDKALKA--KA 756
Db 746 INQOLEELTVSE---VHKQVWENHKQKAFYEPLCEFIKESQTLQIDPDQVTSDDSI 801
Qy 757 EKALVTKKATKNGHILERSINKAVLANNSAIKKAANKKLEKELDLITLVEKGPILAQA 816
Db 802 KKLLESISESKYHAYRLNPSQ-----SSSLIKSEN-----EBIQELKACDEKGVLVKA 851
Qy 817 TMOGVYLLKTPPLPEYVIGLVNVPD 843
Db 852 -----YKFPCLP-----TGTYYD 866

```

RESULT 12 US-09-721-456-221

Sequence 221, Application US/09721456
Patent No. 6617110

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.
Lingner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin B.
Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Crew LLP
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/721,456
FILING DATE: 22-NOV-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015889-00261005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 221:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 221:
US-09-721-456-221

Query Match 3.7%; Score 200; DB 4; Length 872;
Best Local Similarity 20.3%; Pred. No. 1.2e-05;
Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

Qy 14 TTTSVVTYSQAEVYGL-----REESYK-----QEOQASASEDDWFEDEK- 55
Db 17 TNLDFVLQNLBYKSIIEHYKTOQOOIKEDLKLKFKKQDDQNSGNDDEHNSNRQ 76
Qy 56 -----TNSKENSIVDETVSDLFSDGNSNNSSSKTESVSPKQVPPRAKEVTVQESN 108
Db 77 OELLRRVNOIKQOVQILKVGSKVEDNLNL-----EDEN 111
Qy 109 SGNDSKIVPPKQDASKKETLETSTWEA---KDFTRGDTLVGSKG----- 154
Db 112 KKNGLSEQV-----KEQRLRTIEGVKYQNLVFNNDYQDLNESGHRHRRRTDY 164
Qy 155 -INKLSQTHLVPSHADGTOLOVASFAFPDKKTAIAEYTSRLGEGKXPSRLDIDQ 213
Db 165 DREKWEIHS-----DQKNVSIYA---NQKTSICMWLKQYFNKANDHVLNVSIN 211
Qy 214 EIIIDGEIF--NAVQITKLTIPNGKSIQDAPVDNK-----NIAEV- 253
Db 212 RLETEAFYAFDFSGTILTLNNSQYVNIIVNEDNNLCITALLRFLTLRFPNLTNR 270
Qy 254 -----NLPSLETISDYAFAMSLKQYKPEDNKIVYGEIAPFNQIGK----- 297
Db 271 SSYTRNQVNEFKIGELLETFAVVFESHRLQGHVPCFAFQVLANVSSQISVXSQLO 330
Qy 298 LY-----LPRHLIKLAERFAFKSNRIQTVFELSKKVIQGEAS 334

```

```

Db      331 VYSEFTDLKLVDTNKKVQDYFKPLQEPFR-LTHVSCQALPVSATNAENLAVLKKYKHA- 388
QY      335 FQNNLNANVLPGLKEKISEAFNGPDGDEHNNVLTFRTRQONPHQATENTYVNPDK 394
Db      389 -----NLMVLSIP-----TQNFDFYFVNLQHLKLEFGLJEPNILLTKQKL-----E 428
QY      395 SLWRAPDMQYTKLEEDF-TY--QKSVNGFSGNG--LQKVRNKLLEIPKQHNGLTTT 449
Db      429 NLLISTKQSNLKLFLNFTTYVAQETSRKQILKQATTTNKLNNKQOESTPTKQETPS 488
QY      450 EIGDNAPRVDPQSKTLRKYLDEIKLPST--IRKIGAPAFQSNMLKSPAS--EDLJE 504
Db      489 E-STSGMKFFDHSLELEDF-SVNLQATQELYDSLHKLITSTMLKFKLSYKTBMEK 546
QY      505 IXGAP--MNNRIGTL-DLKDKLIKIGDAAFHNNHYAVLPESV-----QET--G 550
Db      547 SKMDTFLDKNLIYETLNKLKRCVSNISNP--HGNI SYELTNKSTFYKFKYLTJNOELQHA 604
QY      551 RSAPFQNGALHLMFIGNKVTIGEMAFSLNKLKESV-----NLSEKQKLTIEVQAPS 602
Db      605 KYTFKQK-----EFQPNVKS--AKIESLSLELEDIDSLCKSIASCNKLQNV----- 650
QY      603 DNALSEVLPNLPNLTREAFKRNHL--KEVKGSSSTLSQITENA-PDQNDGDKRFGKY 658
Db      651 -NIIASLVEYNNIQ--KNFKNPNLLEFKQFQKLNLENVINCILIDQ----- 695
QY      659 VVATNHNHMLADGERFIIDPDLSTMDLEKYLKIEGLDYSTRQTTQ----- 709
Db      696 -----HILNISSELEKNNKIKAFILKRYVLYQY--LDYTKLFKTLQQLPELNOVY 745
QY      710 --TOPREMTTAGKALISKSNIROGEKQK-----FLOEAFELGRVDIDKALA--KA 756
Db      746 INQLEBELTSE--VHQVWENHKQKAFYEPCEFIKSSQTLQIIPDQTVSDDSI 801
QY      757 EKALVTKATKNGHLLERSINKAVLAYNNSAIKKNVKKLEKELDLTLDVBEKGPLAQA 816
Db      802 KKLIESISSKHYHYRLNPSQ-----SSSLIKSEN--EIIQELAKCEKGVLYVA 851
QY      817 TWQGVYLLKTPPLPEYIYGLNVYPD 843
Db      852 -----YKFPPLCLP-----TGTYYD 866

```

RESULT 13

```

US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6547766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowitz, Roman
; APPLICANT: Valsberg, Eugeni
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; PRIOR FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

```

```

Query Match          3.6%; Score 192.5; DB 4; Length 2662;
Best Local Similarity 20.0%; Pred No. 0.0002;
Matches 242; Conservative 170; Mismatches 433; Indels 365; Gaps 59;

```

```

QY      21 YQOEYVGLERESVQEOQOSASEDDWFEEDN--ERKINVSKE NSTVDETS----- 70
Db      507 YNVLVDYQTLTRTEKEMELKLKEKNVDLDEFALREKTKQDQEMQIHLISMLKLVKR 566
QY      71 DLESDQNSNNSKTESVSDKQVPKAPV--TQEASNSDAS--KVEVPKQDAS 125
Db      567 EYVNDQLENEISSKEELREKEDQIKLOEYIDQKLENIKODLSYSLESIEDPKQ----- 622
QY      126 KKETL--ETSTWEAK--DFVTRGDTLVGFSKGIKLSQSHLVLPSHAAGTQLTQYA 180
Db      623 MQQTLFDATVLDKAGSAFIRSENL--LKKMKELATTY--KOMENDIQLYQ-- 673
QY      181 SEAFPPDKKTAJAEYTSRLGNGKPSRLDIDQKEIID--EGELENAY--QLTKLITPQYKS 238
Db      674 -----SOLPAKKMQVDLEKELQSAFNEITKTL----- 701
QY      239 IQGDAPVDKNTAFVNLPESLTISYAFRAHNSIKQVLPDNLKYVIGELAFDPNQIGKL 298
Db      702 -----SLIDGR-----VPRDILC-----NLEEGKITDLOKELNKV 733
QY      299 -----YLPRLILAEAFKSNRIQTV-----FLGSKLVIGEASFQDNMLN 342
Db      734 EENMALREEVILLSELKSLPSEVERLRKEIQDKSEBHLITTEKQDLFSEBVHKEKRYG 793
QY      343 VMLPGLKEKISEAFNG--PDGHHNNQVLTFRTRQONPHQATENTYVNP----- 392
Db      794 LLEEGIKTDLATQSNYSKTDQEFQNKTLHMDFEQKRVLENEENHNOEIVNLSE 853
QY      393 --DKSLWRAPDMQY-----TKMLEDFTYQKNSVTGFSNGK--LQVRNKNLE 438
Db      854 AQKPBSSLGALTELSYTKLOELQKTRVQERLNEHQLEQJENNDSPLQTVREKTLTI 913
QY      439 IPRQHNGL--TIT-----EIGDNAPRVDPQSKTLRKYLDE 472
Db      914 TEKLOQTELEVYKTLQOKKDLKQLOESLQIERDQKSDIHDVYNNNIDQEOQL--RNAL 971
QY      473 EIKLSTIRKIGAFAPQSNLKSFEASEBLEIKGAFNNNRIGTL--DLKDLIKYIGD- 529
Db      972 SKKHQETI--NTLK-----SKISEVSRNLHNEENTGETKQEFQOKRWGIDKK 1017
QY      530 --AAFHINHYAVLPESVQETGRSAFQNGALHLMFIGNKVTIGEMAFSLNKLKESV 586
Db      1018 QDLEKKNQYTLTRDYKQDNELIIOQKIF-----SLIOENELQOMLESV- 1061
QY      587 LSEQKQKLT-----LEVQAFDNALSEVLPNLTQTRB--EAFKRNHL--KEVKGSSSTLS 638
Db      1062 IAEKQKLTDLKENIEMTLENQOELR--LLGDELKKQOETIVAQEKHAIKKEGELSRCTD 1119
QY      639 QITFNAFPQNDQKRGKKVYVTRTNNSHMLADGERFIIDPDKLSTWVDEKYLKING 698
Db      1120 RLA-----EVEEKLEKSOQLOEKQOOLN--VOEBMSQKKNELN 1161
QY      699 LDYSTLRQTTQTOPREMTTAGKALISKSNIROGE-----KQKFLQEAQ--FPLGRVDID 750
Db      1162 LKNEH--KMKELTLHMETBERELAQKNNENVEVKSITKEKVLKELQKSEYTERDHR 1219
QY      751 KALIAAB-KALVTKATKNGHI-----LERSIN-KAVLAYNNSAIKA----- 791
Db      1220 GYIRIEIATVGLQTEKELIAHILKHEQETIDELRRSVEKTAQIINTQDEKSHTKLOE 1279
QY      792 -----ANKRLEK-----ELDILTDLVGEKGPLAQATMGQVYLLKTPPLP 832
Db      1280 ELPVLAHEQCELLPNKVSQETQINNELBLTE-----QSTTKDSTTLARLEME-- 1328
QY      833 EYVIGLVNYPDKSGKLIYAL--DMSDTIGEG-----QKDAYGN 868
Db      1329 --RLRLNEKFOQSGEIKSLTGERDNLKTIKALEVVKDQLEKHIREFTLAKIQEOQSKOE 1386
QY      869 PLIANDV--DNESGYHTLAVALTADYEGLYTIKQILNSSLDKIKAKIROPILAKYHRL--GIFQA 926
Db      1387 QSLNKEKQDN-----TKIVSEWQFKPKD--SALURIE--IEMIGLSK--FLQESHDE 1435

```

QY 927 IRMAAABDLRLPTPKGYLNEVPNRYKKONEKULKPVDTYKPIFNKALPNEKYDGRAA 986
 Db 1436 MKSAKEDDL-----ORLOEVLQSESDQKEMIK-----EIVAKHLE-----EELK 1479
 QY 987 KGH---NINATNNSVAVTPIRSE-----QOLHKSQSDVNI.POT 1022
 Db 1480 VAHCCLEQOEETIELNRLVNLSEKETESTIOKOLEAINDKLQNKIOETIYEKOEOLNIKOI 1539
 QY 1023 SSKNNFIYEI 1032
 Db 1540 SEVOENNEL 1549

RESULT 14
 US-09-107-532A-3959
 Sequence 3959, Application US/09107532A
 Patent No. 6583275
 GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 7310
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354
 COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/107,532A
 FILING DATE: 30-Jun-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/085,598
 FILING DATE: 14 May 1998
 APPLICATION NUMBER: 60/051,571
 FILING DATE: July 2, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Arinello, Pamela Denise
 REGISTRATION NUMBER: 40,489
 REFERENCE/DOCKET NUMBER: GTC-012
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-5007
 TELEFAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 3959:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1221 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (B) LOCATION 1...1221
 SEQUENCE DESCRIPTION: SEQ ID NO: 3959:
 US-09-107-532A-3959

Query Match 3.5%; Score 189; DB 4; Length 1221;
 Best Local Similarity 19.0%; Pred. No. 0.00012;
 Matches 228; Conservative 14; Mismatches 403; Indels 394; Gaps 61;

QY 4 KHLKTLAL--TTVSVTVYSQEVY---GLRESYKQOQOSASDDWFEEDNERKTNS 59
 Db 14 KHLKTLAL--TTVSVTVYSQEVY---GLRESYKQOQOSASDDWFEEDNERKTNS 59
 QY 60 KENS---TVDETVSDFLSDGNS--NNSSKTESV---VSDIKQ--VPAKKEVTOEASN 108

Db 74 GAGLMTVVVVDNDPEFLLEBSPLKNOQENTGAVEGVBEKREKADKANEEKVKNPDDI 133
 QY 109 SSSDASKVEY---PK-----ODTASKEETLETSTWPAKQFV--TRGDTLVGFEKSGNK 157
 Db 134 DYNDISSGLANDPKRPFLSEBELNKLQTKNTEFSEALMKTR-----GYPSORILE 188
 QY 158 LSQTSILY-----LPSHADGTQTOVASFAFPDKKTAIAEYTSRLGENG-----KPS 206
 Db 189 LSNMNDVPRKAVSRHVALVSNRAQVSETRINPPNGYPMVYILIR--SNQSYTENVAI 246
 QY 207 RLIDQKEII-----DEGEIFNAYQLTKLTPNGYSIG----- 240
 Db 247 RRTIDGKEAFCLQAGVQANTTYTGSTGYSNSQREMISSILAYKELGGWANNADAKOR 306
 QY 241 ---ODAFVQNKINIAEVNLPESLETISDY-----AFAMSLKQVKKPDLNKVI 284
 Db 307 WATQMLWENSN---EKPKSLTVPMWKGATKMKNDLVAANALRYTKMDSTSKTV 362
 QY 285 --GELAF-----DNQI-----GKLYLPRHLIKLAERAFKSNRIQVEFL----- 323
 Db 363 DEGDKATFKVNTYTGDNHLVKSQSGGKATISNGTLVDTGATGNKI--TYDMIKGSH 421
 QY 324 -----GSKLVIGESAFQDNNIR--NWMLPDGLE--KIESE-----AF 357
 Db 422 SDAPFTTWSNGSYQKLVGTSFYISTSVYVVRKGTVEVLKVDSETKKPLSGATPRFSY 481
 QY 358 TGNPDGEHNNQVVLRTRGONPHOLATE---NTYVNDPKSLMRATPMDYTKMLEEDF 413
 Db 482 SGKTKD-----VTTDSNGKAKLTLEKAGTVK-----YKEIKAPN 517
 QY 414 TYQKN---SVYFSKSGLOKAPRNQOLBIPKOHNGITTEIGD--NAFRNDF----- 461
 Db 518 GYQLDSEFSFITYKEQNVYTTTNTK---KSTGVSVEIEKIGDLGGLPNVVFITYNS 572
 QY 462 QSKTLR-----KYDLEIKLPS--TIKIKGAPAFQSNMK--SEASEDELEIKEGA 509
 Db 573 DNKVVADNKLTKSNGKIKVDELQPKRYAVEKQGTGYDPDGKKVFEISQDPPFRK--- 629
 QY 510 FNNNRIGTLDLQDKLIKIGDAAFHINHYAIVLPESVQIGHSAPFQNGALHIMFGNKV 569
 Db 630 -----PAKVVTNIEVSEISEKVSQMNNSQVLKNN----- 659
 QY 570 KILGEAFSLNLESNVL--SEKQKLTIEVQAFSNALSEVYLPPNLOTIREAPKRNH 627
 Db 660 WSADEPDVYQYDITTGNIKKSGERNVQSPFIIGFYDN-----KNVDSKSEVY--- 707
 QY 628 LKEVKGSSSTLSQITFNAFPQNDGDKRFQKVVVRTHNNSHMLADGERFIIDDKLSSTNV 687
 Db 708 ---VGSNTVT---NSPDIINDT-----NNGKVIKAKTISVLTNNDFYKSY 747
 QY 688 DLKVKIKTEGLDYSTLRQTOGTQPREMTTAKALLSKSNLRQGEKQKLEQAQFLGRV 747
 Db 748 NLRITMKI-----KSKTLNTEKQKNL----- 769
 QY 748 DLDDKALAKERKALVTKKATKNGHLERSINKAVLAYNNSAIKAVKRL--EKSLDILDL 806
 Db 770 ---TIVKSKVTVSGKSKSA---SSNDVETVYS--RKVTINHIDEXDKHLKXD 817
 QY 807 VEGK--GPLAQATWQGVYLLKTPPLPEYVYIGLVNVPFKSGKLIALALMSDTIGEGOK 864
 Db 818 IDYKTDGETYE---YKPRITL-----FDKQNNNT---KSVITHKQKD 854
 QY 865 AYGNPILAVNDENEGHTLAVATLADYGLYKIDILNSSLDKIKAIROIPIAKYRLGIF 924
 Db 855 --GNDIVL---NTFYHI-----PVLAVVDRI---QIDTAR----- 882
 QY 925 QAIRMAAABDLRLPKTKRGYL--NEVPNRYKKONEKULKPVDTYKPIFNKALPNEKYD 983
 Db 883 -AVKA-----GHPTKIEISKEEYKELEKTVFKYKIID--IDNNKVVYD 925
 QY 984 PAKGHNINATNNSVAVTPIRSEQOLHKSQSDVNI.POTSXNNFIYE-----ILGYVS 1037

Db 926 ENFKYKERNSHSLDPTLEYITKKNKINYS-VDILLVNDPDKKFEETETKDLRTIGTYS 983

RESULT 15

US-08-592-126-148

Sequence 148, Application US/08592126

Patent No. 5821091

GENERAL INFORMATION:

APPLICANT: Gregory Dolganov

TITLE OF INVENTION: Transcripts Encoding Immunomodulatory

TITLE OF INVENTION: Polypeptides

NUMBER OF SEQUENCES: 151

CORRESPONDENCE ADDRESS:

ADDRESSEE: Denlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/592,126

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles K.

REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 4600-0111

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-0880

TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 148:

SEQUENCE CHARACTERISTICS:

LENGTH: 1312 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYDROTHERMAL: NO

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: Rad50. Pro-translation of SEQ ID NO:54

US-08-592-126-148

Query Match 3.5%; Score 188; DB 2; Length 1312;

Best Local Similarity 19.2%; Pred. No. 0.0015;

Matches 220; Conservative 185; Mismatches 446; Indels 296; Gaps 50;

Db 18 VVTYSQEVYGLH-REESVQKQOTQASSEDWFEEDNERKTNVSKENSTVDVTSDF--S 74
 Db 241 VKSENELDLKMLKEIEHLNLSKIMKLDNFKALDSKKMKDKNSLEKMKVQGT 300
 QY 75 DGNSSNSSKTESVSPKQVPAKPEVTOEASNSSNDASKEVVPKODTASKEETEST 134
 Db 301 DEOUNDLVYHQRVY--REKEREKLVDCHELEKLNKESRLINQESSELVVEQRLQQA 357
 QY 135 WEAKDFVRGDTLVGFSGSGINKLSQTSHTLVLPFHADGTOITVQASAFPPDKTAAE 194
 Db 358 DRHOEHIPARDSL- 395
 QY 195 YTSRLGNGKPSRLDIDQKEIIDGELFNAYQLTFLTPNGYKSIGQDAFVNDKNIAEVN 254
 Db 396 FHKLVRR- 428
 QY 255 LPE- 308
 Db 429 IDELRDKKTGRIIEKSKKQNEL-KVVKY--ELQQLGSSDRILIEDDELIAE 485
 QY 309 ERAFEKSNRIGTVEFISGLKTYI--GEASFDNNLRNMLPDGLEKIESEAFNPGDEH 365
 Db 486 RELSKAKGNSNVETL--KMEYISLONEKADIDRITR- - - - - - - - - - - - - - - 527

QY 366 YNQVAVLRTGQPHQALATENTVYNDPKSIM--RATPDMDY--TKMLEEDFTYQKNSVT 421
 Db 528 LNHHTTTTQWEMLTQKADQDEQIRKIKSRHSDELTSLGYPNNKQLEDMILSKSKEI 587
 QY 422 GFS- 468
 Db 588 NOTDRKLAKLNKEALSSQONHINNEIKRREBSLSTE--DKLFDVCGSDPFES- - - - - 640
 QY 469 YDL- 507
 Db 641 -DLRLKEIEKSKQAMLAGATAVYSQFITQTDENQSCCPVQGRFQTEABEQV- - - 697
 QY 508 GAFNNNRIGTDLQDKLIKGDAAFHNNHYAIVLPESVQIEGSAFRQNALHIM- - - 563
 Db 698 - 737
 QY 564 -FIGNKVKTIGEMAFLSNKLESVNLSEBQKLTIEVOAFSDNALSEVLPNNLTIREA 622
 Db 738 SIIDLKEKEIFE--LNKQLQNVNRDIOQLKNDIBEQ--ETILGTIMP- - - - - 784
 QY 623 FK- 664
 Db 785 AKVCLTDTVTIMERFQMELEKDYERKIAQQAQKQIDLDRTVQVNOQEKQKHLDITVSS 844
 QY 665 - 717
 Db 845 KIELNRKLIQDQOQI--OHLKSTNELKS--EKLQISTVLQRRQQLQEEQVETLST 896
 QY 718 AGKALSKSNLRQGEKO- 768
 Db 897 EVQSLYNE--IKQAKEQVSLPTELEKFOQKEKELINKKTSNKLQ--DKLNDIKKAYKN 953
 QY 769 GHLERSINKAVLAYNNSAIKANVKELEKLDLITLVESKGPLAQ--ATWQGVYLLK 826
 Db 954 IHGYMKDIEN-YIDGKODYKQKQETELNKVIAQLSECEKHKEKINEDMLRMQDIDTQK 1012
 QY 827 TPLDLPYVYIGLVNYPFKSGKLIYALDMSDTIGGQGDAGNPILANDDENEGHTTAAVA 886
 Db 1013 - 1043
 QY 887 TLADYEGLYIK--DILNSSLDKIKATROIPLAKYHRLGIFQAIRNAAAEADRLLPKTPK 943
 Db 1044 EMGQGVQLQKMSHQKLEENIDINKRNHNLALGR- - - - - - - - - - - - - - - 1079
 QY 944 GYLANEVNRYKKQKQEKVYDYK- 992
 Db 1080 GYEEELIHFKEELREPOFDAEEKYRREMIWRTTELVNKDLDIYKTLQAIWKFSMK 1139
 QY 993 AETNSVAATPIRSE--QQLH- 1042
 Db 1140 MEELINKIIRLUMSTYTGQDIETIETIRSDADENVSASDKRNINRYVLMKGTALDMRG 1199
 QY 1043 LVTAAGK 1049
 Db 1200 RCGAGK 1206

Search completed: April 27, 2004, 11:18:08
 Job time : 28 secs